The international journal of science / 2 February 2023

nature



FUNKY WORM

220-million-year-old caecilian fossil refines the rhythm of amphibian evolution

COVID conundrum
How vaccine makers hope to outwit coronavirus evolution

Health and heredity
Four steps to unlock Africa's potential in medical genomics

Impromptu action
Dopamine helps shape spontaneous behaviour in mice

VISIT…

LANZAROTE Caliente.com

Nature.2023.02.04

[Sat, 04 Feb 2023]

- **This Week**
- **News in Focus**
- **Books & Arts**
- **Opinion**
- **Work**
- **Research**
- **Amendments & Corrections**

This Week

- **Is science really getting less disruptive — and does it matter if it is?** [25 Jan 2023]
  • A study suggesting papers and patents that change the course of science are becoming less dominant is prompting soul-searching — and lively debate about why, and what to do about it.
- **Why we have nothing to fear from the decolonization of mathematics** [31 Jan 2023]
  • Maths made the modern world — and everyone stands to gain from the acknowledgment that the world made maths.
- **Academia's culture of overwork almost broke me, so I'm working to undo it** [31 Jan 2023]
  • For young immigrant women like me, the pressures of early career research are even greater than for most. But it doesn't have to be that way.

- EDITORIAL
- 25 January 2023

Is science really getting less disruptive — and does it matter if it is?

A study suggesting papers and patents that change the course of science are becoming less dominant is prompting soul-searching — and lively debate about why, and what to do about it.

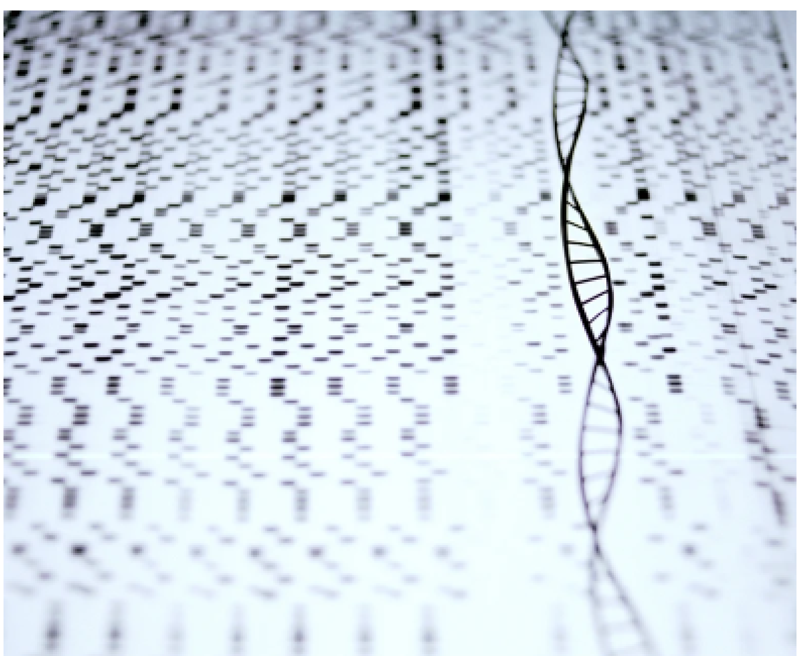


The discovery by Michel Mayor (right) and Didier Queloz of a planet orbiting a Sun-like star launched the search for exoplanets.Credit: Karl Schoendorfer/Shutterstock

The influential twentieth-century physicist and philosopher Thomas Kuhn was instrumental in formulating the term ‘paradigm shift’ to characterize how unexpected evidence can set research fields off in new directions. A paper published in *Nature* earlier this month by the social scientists Michael Park, Erin Leahey and Russell Funk has prompted lively debate by suggesting that the proportion of disruptive papers and patents has been decreasing over time[1].

By analysing more than 60 years of data from bibliometric and patent databases, the authors conclude that it is less likely now than in the mid-twentieth century that any one paper or patent will be ‘highly disruptive’ — that is, that it will change the course of an entire scientific field. Although the number of new papers and patents the researchers classified as disruptive stayed broadly the same over the period they studied — from 1945 to 2010 — the explosion in research articles, patents and funding in that time means that disruptive science’s share of publishing and patenting has been dropping.

Much of the reaction has involved soul-searching about the implications for science if innovation is slowing down, as well as questions about the nature of the modern scientific enterprise itself. This, in turn, is prompting more questions that could become the subject of further analysis.



'Disruptive' science has declined — and no one knows why

The study uses a number of measures of disruptiveness. The one that has attracted perhaps the most attention is called the CD index, which is based on citations. As the authors write, "if a paper or patent is disruptive, the subsequent work that cites it is less likely to also cite its predecessors", whereas "if a paper or patent is consolidating, subsequent work that cites it is also more likely to cite its predecessors". In other words, with more consolidation, the same previously disruptive papers continue to be cited.

Single papers do have the potential to disrupt or create fields. One of the best-known examples is James Watson and Francis Crick's model of DNA from 1953, created with the help of Rosalind Franklin's ground-breaking X-ray crystallography work[2,3]. Another is the 1995 discovery by Michel Mayor and Didier Queloz of a planet orbiting a Sun-like star[4] that launched the field of searching for exoplanets.

But new directions also arise from many studies reporting long-running research efforts. Gravitational waves are one example. Much as the paper from the LIGO collaboration reporting the first direct detection of a gravitational wave[5] is itself highly cited, subsequent work has continued to cite work that led up to it. Researchers cite studies for different reasons, and not only to acknowledge previously important work that is being built on. Park and his colleagues do control for some of these things to better compare disruptiveness today with that several decades ago.



[The science that’s never been cited]

For this Editorial, *Nature* spoke to a number of scholars who study science and innovation. The paper by Park and his colleagues[1], they say, builds on a pattern identified elsewhere in the specialist literature[6,7], and some are worried by the findings’ implications. Science and innovation are drivers of both growth and productivity, and declining disruptiveness could be linked to the sluggish productivity and economic growth being seen in many parts of the world.

Others argue that a decline in the fraction of disruptive science shouldn’t cause concern if the absolute number of disruptive studies has remained relatively constant over time. If a greater proportion of publications are consolidating, that could just reflect the current scientific situation: in many disciplines, the fundamentals are agreed on, so most further advances will be incremental, rather than disruptive.

It is also possible that researchers are recording more incremental steps in formal research papers than before, rather than waiting to report something of greater significance. No doubt scholars will further analyse the importance of these findings using qualitative empirical approaches, such as detailed interviews and observations that capture researchers’ own experiences in individual fields, as the sociologist Harry Collins has done from within the LIGO team.

Growing division

Another reason why the study by Park and his colleagues has created such resonance is that it plays into wider concerns about how science is organized. One of these is whether the division of science into ever-narrower units of knowledge is detrimental to the discovery of new paths. Critics also point to publication incentives and metrics-driven research evaluations, which steer scientific study away from risk-taking as funders, researchers and institutions take the safe option to keep the grant–publication–citation wheel turning.



Bibliometrics: Is your most cited work, your best?

This periodically leads to calls to either incentivize or directly fund more high-risk, high-reward research, and initiatives such as the United Kingdom's Advanced Research and Invention Agency. This is modelled on the US Defense Advanced Research Projects Agency, which was founded in 1958, so the search for disruptive innovation is not new.

But it is also possible that science's knowledge and publication overload is not specifically a research problem. The lack of space to think in the face of an information deluge is increasingly apparent across many sectors of society. Some in innovation studies think that artificial intelligence could

help, by sifting and sorting information in meaningful and beneficial ways: aiding researchers in summarizing cutting-edge knowledge in a discipline[8], for example, or identifying which research projects have the potential for breakthroughs[9]. If used appropriately, such technological disruption has the potential to free up more time for scientists to progress their fields — disruptively or otherwise.

Asking questions about the nature of science and reflecting on the answers can only be a good thing. The work by Park and his colleagues must continue to be built on, using both quantitative and qualitative methods, down to the level of individual fields. This will help us to understand in more detail how and why science is changing, and where we want it to lead. The end result could be disruption or consolidation — or even a paradigm shift.

Nature **614**, 7-8 (2023)

doi: https://doi.org/10.1038/d41586-023-00183-1

References

1. Park, M., Leahey, E. & Funk, R. J. *Nature* **613**, 138–144 (2023).
2. Watson, J. D. & Crick, F. H. *C. Nature* **171**, 737–738 (1953).
3. Franklin, R. E. & Gosling, R. G. *Nature* **171**, 740–741 (1953).
4. Mayor, M. & Queloz, D. *Nature* **378**, 355–359 (1995).
5. Abbott, B. P. *et al. Phys. Rev. Lett.* **116**, 061102 (2016).
6. Bloom, N., Jones, C. I., Van Reenen, J. & Webb, M. *Am. Econ. Rev.* **110**, 1104–1144 (2020).
7. Chu, J. S. G. & Evans, J. A. *Proc. Natl Acad. Sci. USA* **118**, e2021636118 (2021).

8. Wagner, G., Lukyanenko, R. & Paré, G. *J. Inf. Technol.* **37**, 209–226 (2022).

9. Weis, J. W. & Jacobson, J. M. *Nature Biotechnol.* **39**, 1300–1307 (2021).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00183-1

- EDITORIAL
- 31 January 2023

Why we have nothing to fear from the decolonization of mathematics

Maths made the modern world — and everyone stands to gain from the acknowledgment that the world made maths.

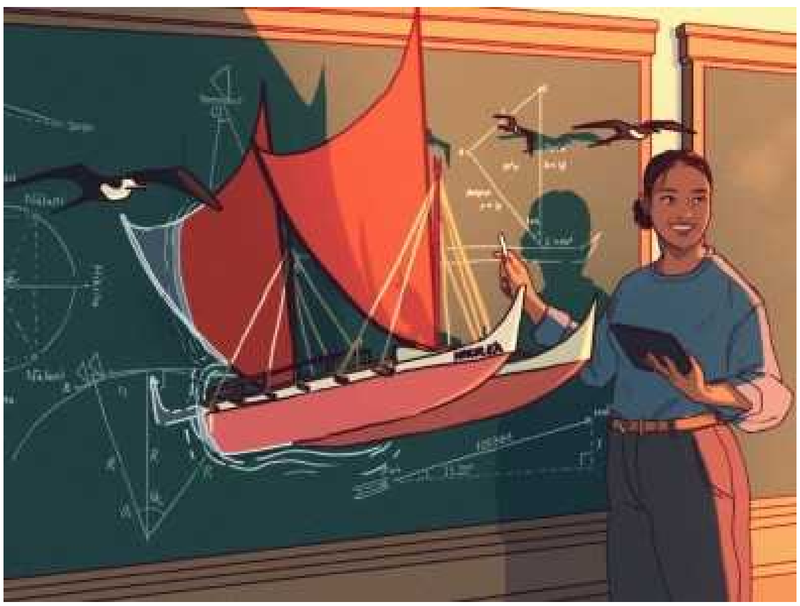


The Bakhshali manuscript was found in 1881 around 20 kilometres from the site of an ancient Buddhist monastery in what is now Pakistan.Credit: Dinendra Haria/Alamy

What's the point of decolonizing mathematics? This is one of the questions being asked in response to the international decolonizing movement, through which university faculty members and students are exploring the

contributions that people from many cultures have made to the story of different research fields.

Such questions have long been central to the study of science's history, but not its teaching. Efforts to address them include works such as maths historian George Joseph's book *The Crest of the Peacock: Non-European Roots of Mathematics* (1991) and the ongoing encyclopaedic series Science Across Cultures, edited by Helaine Selin. But there's been pushback, too, as interest in decolonization has mushroomed in many parts of the world. The debate on decolonizing mathematics is explored in the latest instalment of our series on decolonizing science.



Charting a course to make maths truly universal

Some researchers are concerned that the decolonization movement politicizes universities and restricts academic freedom. One common argument is that decolonization is irrelevant to the practice of mathematics: the solution to a quadratic equation doesn't, after all, depend on a mathematician's identity or protected characteristics.

In fact, such questions reprise aspects of an older, more academically focused debate on whether — or to what extent — scientific knowledge is socially constructed. What is known as pure mathematics is a case in point. Algebra, the method of representing problems in mathematical form, is often

taught as a collection of rules to memorize and practise, described using abstract symbols such as *x* and *y*. But it was not always so.

The word 'algebra' is a contraction of the name of a ninth-century Arabic text, *al-Kitāb al-mukhtaṣar fi ḥisāb al-jabr wa'l-muqābala* (The Compendious Book on Calculation by Completion and Balancing), written by Muḥammad ibn Mūsā al-Khwārizmī, who was born in about ad 780. Al-Khwārizmī was an astronomer, cartographer and mathematician working in Baghdad, in what is now Iraq, during Islam's imperial era. He wrote the book as a kind of public service, to help people work out everyday problems, such as what they were entitled to in inheritance or how much they owed in tax. The text includes worked examples and uses words instead of symbols, along with visual and geometric techniques to make solving problems easier.



[Imperialism's long shadow: the UK universities grappling with a colonial past]

Mathematicians from Europe, particularly Fibonacci (born in Italy in around 1170), came across these ideas while travelling in Arabic-speaking countries and helped to extend their geographical reach by presenting them in Latin. Chief among the concepts Fibonacci helped to introduce to Europe was the Indo-Arabic system of numerals we use today. Because many original Arabic texts (including Al-Khwārizmī's algebra book) were later lost, many

ideas that originated in the Middle East later found their way back to the region through translations of these Latin texts.

Al-Khwārizmī refers to Indian numerals to acknowledge earlier sources for his methods. The only surviving example of a potential source of this type is the Bakhshali manuscript, a collection of some 70 ‘pages’ of mathematics written on birch-tree bark in a form of Sanskrit. The manuscript has been dated to as early as the third or fourth century ad and is now stored at the Bodleian Libraries at the University of Oxford, UK. It is regarded as the oldest surviving text that uses the concept of zero (represented by a dot). Even today, numbers written in Arabic use a dot to denote zero.

This manuscript also contains descriptions of what would have been everyday mathematical problems at the time, and rules for how to solve them. Among these are linear equations, quadratic equations and means of finding the square roots of numbers. Like Al-Khwārizmī, the unknown writer describes equations using words.

So, to answer the question: what’s the point of decolonizing mathematics? It is so we can get a more accurate picture of the subject’s origins and development, and the variety of problems it helps to solve. Decolonization shows that the roots of discovery and invention are shared between many world cultures, which can be particularly empowering for people from historically marginalized groups. Decolonizing science is the antidote to exceptionalism, the idea that any single culture or civilization possessed special abilities in advancing science.

The last words must go to the guest editors of *Nature*’s special issue on racism, published in October. “It is so important for science curricula, research and academic spaces to go through decolonization processes. These are not political or ideological acts, but part of science itself — an example of science’s self-correcting mechanism in the pursuit of truth.”

Nature **614**, 8 (2023)

doi: https://doi.org/10.1038/d41586-023-00240-9

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00240-9

- WORLD VIEW
- 31 January 2023

Academia’s culture of overwork almost broke me, so I’m working to undo it



For young immigrant women like me, the pressures of early career research are even greater than for most. But it doesn’t have to be that way.

- Natalia Ingebretsen Kucirkova

I left Slovakia to study in the United Kingdom at the age of 18, with a small suitcase and big hopes. I wanted to take advantage of opportunities that my parents had only been able to dream of growing up under a communist regime. But never in my wildest dreams did I think I would become a professor before the age of 35.

I got here by elbow grease — and long hours. I worked feverishly to get funded PhD and postdoctoral positions. The stiff competition for grants and positions at high- ranking universities helped me to crystallize my research ideas in my field of early education and child development. But to stand out from my peers, I doubled my working hours. It wasn't uncommon for me to work 12 hours a day, 7 days a week, for months at a time.

More hours meant more results. Seeing my studies on digital reading translated into apps for children or family websites motivated me to do more. I put my work on a pedestal, often at the cost of my health and social life.

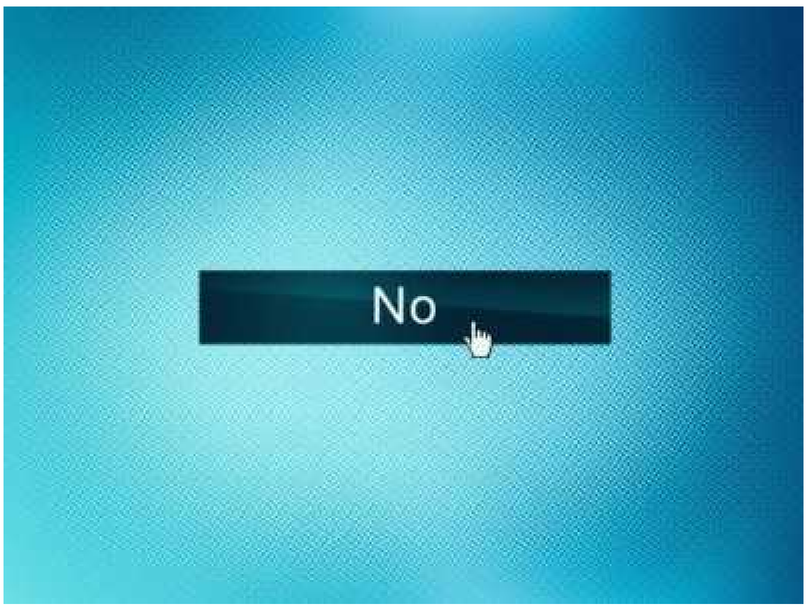


How burnout and imposter syndrome blight scientific careers

I thought I was the exception. But when I read an article containing interviews with five successful female psychology researchers (P. Alexander *et al. Educ. Psychol. Rev.* **33**, 763–795; 2021), I realized that this is the norm for top-performing academics. I greatly admire the interviewees and share their passion for their work. But I now realize that, by hiding behind passion, I was excusing my contribution to a toxic burnout culture in research. And for me and many others like me — female, immigrant, non-native English speakers — the pressures are even greater. It's time to speak out.

I made the greatest sacrifices during my years on temporary postdoc and lectureship contracts, when not publishing an extra paper could have cost me the grant I needed to secure next year's salary. A mentor told me that the passport to academia is publications, so I filled every spare moment with writing. A doctor told me ice would ease my permanent carpal tunnel syndrome, so I typed wearing iced wrist splints.

Not being a native English speaker, I had to put in extra hours for each paper. The fear of being misunderstood by using the wrong word added to the stress of conference presentations and translated into regular pounding headaches and fatigue, which I still experience.



[Why four scientists spent a year saying no]

The pressure to perform sucked me into a negative spiral. When I felt stressed, I doubted myself, feared saying 'no', overcompensated by saying 'yes' to extra tasks, and became more stressed. I cut back on spending time with friends and on sleep. My then-boyfriend told me I was married to my computer and cut our holiday short when he saw me typing a paper on the beach. The ticket inspector on the late-night commuter train knew me by name because I regularly overslept my stop. When I had a bout of autoimmune illness, my family was not surprised.

I see now that my choice to work hard was fuelled not only by my love for the work but also by systemic factors. Studies show that the risk of burnout is higher among young researchers (A. Boone *et al. Front. Psychol.* **13**, 839728; 2022), as well as among female academics from marginalized groups, because there is greater pressure to perform. Although that includes me, I can't speak to the even greater pressures that affect many young women, among them those from minority racial groups, those juggling motherhood with early-career research, those from the LGBTQ community, and scientists from countries where there is extreme gender discrimination or violent conflict.

Through a combination of hard work and luck, I got a permanent position early in my career. But the workload has only grown heavier as I have climbed the career ladder, with increasing requests for mentoring, article and grant reviews, departmental duties, committee memberships, and voluntary contributions of time and expertise to professional societies. The costs of making a mistake are also higher: if my lower performance delays a large grant, that can jeopardize several people's salaries.



Saying 'no' in science isn't enough

But my survival anxiety has lessened. Starting a family and moving to Norway, a country known for a better work–life balance than the United Kingdom, helped. Beginning to take my childhood hobby of writing poetry

seriously was the best thing I ever did for my mental health. And I have learnt to manage my calendar better, blocking out time to write and not feeling guilty for setting out-of-office replies.

The extreme workload of my early career was unhealthy for me, and it's unhealthy for others. I want to undo my contribution to this toxic culture of overwork, especially for groups that are disproportionately affected.

I see it as my responsibility to promote definitions of academic success that are not tied to extreme working hours. In the book *Inspirational Women in Academia: Supporting Careers and Improving Minority Representation* (2023), my colleague Loleta Fahad, who is head of career development at University College London, and I interview female academics and administrators. We openly share where we failed, what we wish we had known when we started working at a university and what those in power could do to address systemic discrimination.

The pursuit of science lends itself to fervour: there is no ceiling to knowledge, and the discovery process can be all-consuming. But being passionate about our work should not be equated with working extreme hours. And it should not put extra pressure on women from marginalized backgrounds.

Nature **614**, 9 (2023)

doi: https://doi.org/10.1038/d41586-023-00241-8

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00241-8

News in Focus

- **What time is it on the Moon?** [24 Jan 2023]
  • Satellite navigation systems for lunar settlements will require local atomic clocks. Scientists are working out what time they will keep.
- **The next worrisome coronavirus variant could come from China — will it get detected?** [24 Jan 2023]
  • Genomic surveillance is crucial for tracking the next 'variant of concern', but many countries are winding back their monitoring.
- **Scientists petition UCLA to reverse ecologist's suspension** [24 Jan 2023]
  • Sanctions on Priyanga Amarasekare have baffled supporters, who think they are retaliation for speaking out against discrimination.
- **Bird flu outbreak in mink sparks concern about spread in people** [24 Jan 2023]
  • A variant of H5N1 influenza that can spread between mammals could pose an increased risk to people and wild animals.
- **Massive health-record review links viral illnesses to brain disease** [23 Jan 2023]
  • Study ties common viruses such as flu to Alzheimer's and other conditions — but the analysis has limitations, researchers warn.
- **Has Earth's inner core stopped its strange spin?** [23 Jan 2023]
  • Earthquake data hint that the inner core stopped rotating faster than the rest of the planet in 2009, but not all researchers agree.
- **Will a new wave of RSV vaccines stop the dangerous virus?** [27 Jan 2023]
  • Pfizer, GSK and Moderna are ahead in the race to produce vaccines against respiratory syncytial virus — what will be their impact?
- **The next generation of coronavirus vaccines: a graphical guide** [01 Feb 2023]
  • New technologies might provide more potent or broader immunity — but will have to fight for market share.

- NEWS
- 24 January 2023
- Correction 02 February 2023

What time is it on the Moon?

Satellite navigation systems for lunar settlements will require local atomic clocks. Scientists are working out what time they will keep.

- Elizabeth Gibney



Scientists and space agencies have not yet agreed how to define lunar time.Credit: NASA Goddard

The coming decade will see a resurgence in lunar exploration — including dozens of missions and plans to establish permanent bases on the Moon. The endeavours pose myriad challenges. Among them is a subtle, but

fundamental, question that metrologists worldwide are working to answer: what time is it on the Moon?

“We’re just starting to lay this out,” says Cheryl Gramling, an aerospace engineer who leads the position, navigation and timing team at NASA’s Goddard Space Flight Center in Greenbelt, Maryland.



[The $93-billion plan to put astronauts back on the Moon]

The Moon doesn’t currently have an independent time. Each lunar mission uses its own timescale that is linked, through its handlers on Earth, to coordinated universal time, or utc — the standard against which the planet’s clocks are set. But this method is relatively imprecise and spacecraft exploring the Moon don’t synchronize the time with each other. The approach works when the Moon hosts a handful of independent missions, but it will be a problem when there are multiple craft working together. Space agencies will also want to track them using satellite navigation, which relies on precise timing signals.

It’s not obvious what form a universal lunar time would take. Clocks on Earth and the Moon naturally tick at different speeds, because of the differing gravitational fields of the two bodies. Official lunar time could be based on a clock system designed to synchronize with utc, or it could be independent of Earth time.

Representatives of space agencies and academic organizations worldwide met in November 2022 to start drafting recommendations on how to define lunar time at the European Space Research and Technology Centre of the European Space Agency (ESA) in Noordwijk, the Netherlands.

Decisions must be made soon, says Patrizia Tavella, who leads the time department at the International Bureau of Weights and Measures in Sèvres, France. If an official lunar time is not established, space agencies and private companies will come up with their own solutions, she says. “This is why we want to raise an alert now, saying let’s work together to take a common decision.”

Tracking satellites

The most pressing need for lunar time comes from plans to create a dedicated global satellite navigation system (GNSS) for the Moon, similar to how GPS and other satellite navigation networks enable precise location tracking on Earth. Space agencies plan to install this lunar GNSS from around 2030. ESA approved a lunar satellite navigation project called Moonlight at its ministerial council meeting on 22 and 23 November 2022 in Paris, and NASA established a similar project, called Lunar Communications Relay and Navigation Systems, last January.

Until now, Moon missions have pinpointed their locations using radio signals sent to large antennas on Earth at scheduled times. But with dozens of missions planned, “there’s just not enough resources to cover everybody”, says Joel Parker, an engineer who works on lunar navigation at the Goddard Center.

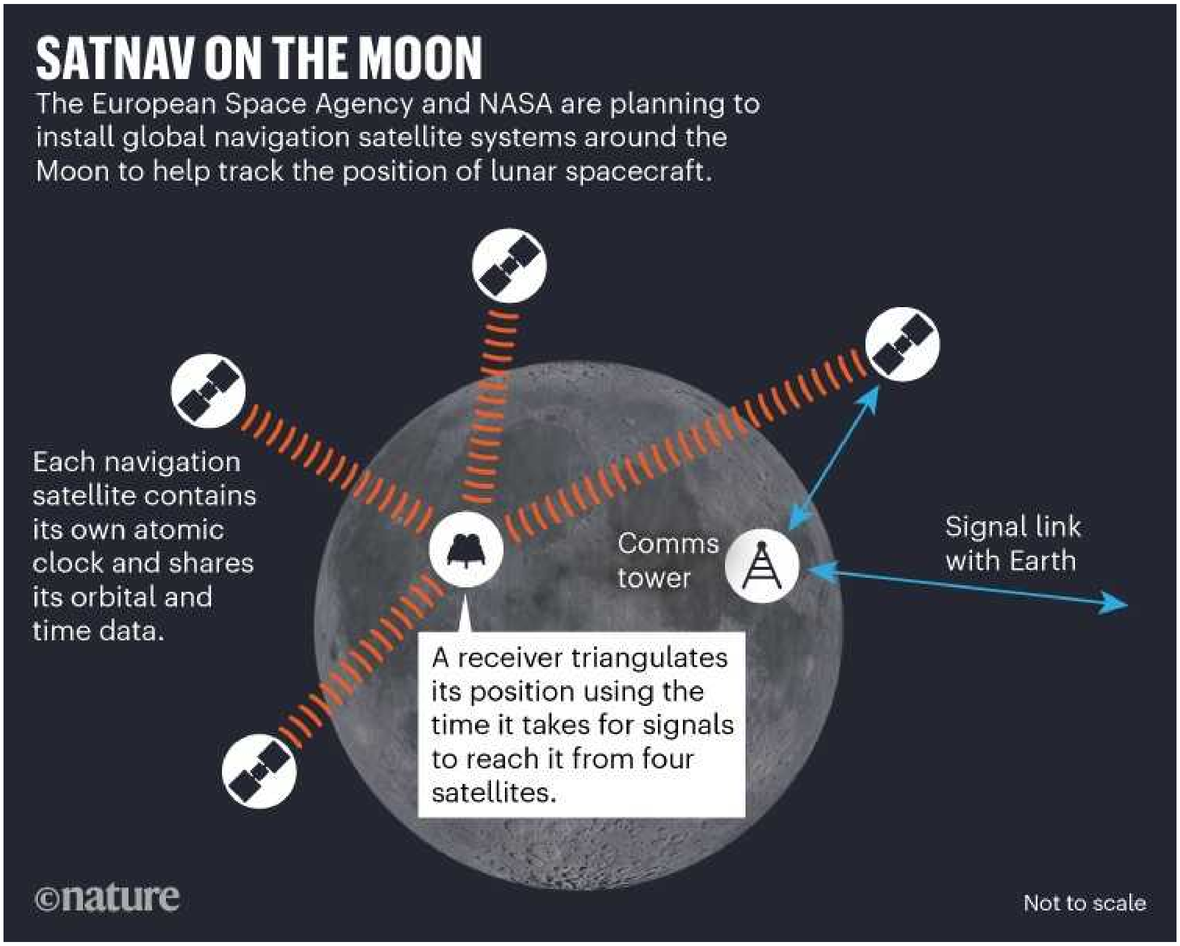


Image credit: NASA

As a first step, from 2024, ESA and NASA will trial deriving positions on the Moon using faint satellite navigation signals from Earth-based craft. Next, the lunar GNSS projects plan to place dedicated satellites around the Moon, each containing their own atomic clock (see ‘Satnav on the Moon’). A receiver, for example on the Moon’s surface, will then triangulate its position using the time it takes for satellite signals to reach it. ESA has planned an initial constellation of four spacecraft that would cover navigation at the lunar south pole, which harbours much of the Moon’s water and is an important target for exploration, says Jörg Hahn, an engineer working on ESA’s Moonlight project.

Moon missions will also need an official lunar time to cooperate and communicate, says Hahn. “All this has to trace to one kind of a time reference, otherwise you have chaos and things do not work together.”

Another open question, says Hahn, is whether astronauts would use universal lunar time everywhere on the Moon. Although lunar time would remain the official timescale, its users might, as on Earth, want to offset it in time zones that link to the Sun's position in the sky. This is less a question for metrologists and more one of convention. "When somebody really lives there on the Moon, I think it makes sense," he says.

Characterizing time

Defining lunar time is not simple. Although the definition of the second is the same everywhere, the general theory of relativity dictates that clocks tick slower in stronger gravitational fields. The Moon's gravitational pull is weaker than Earth's, meaning that, to an observer on Earth, a lunar clock would run faster than an Earth one. Gramling estimates that a lunar clock would gain about 56 microseconds over 24 hours. Compared with one on Earth, a clock's speed would also subtly change depending on its position on the lunar surface, because of the Moon's rotation, says Tavella. "This is a paradise for experts in relativity, because you have to take into account so many things," she adds.

Defining a lunar standard, with which all clocks are compared, will involve installing at least three master clocks that tick at the Moon's natural pace, and whose output is combined by an algorithm to generate a more accurate virtual timepiece (see 'How to build a Moon clock').

What happens then depends on which option metrologists choose. They might decide to base lunar time on utc. In that case, this virtual lunar time would be synchronized regularly with terrestrial utc. Between the check-ins, the lunar master clocks would keep marking time until the next synchronization. This has the advantage of being simple for users back on Earth to interact with.

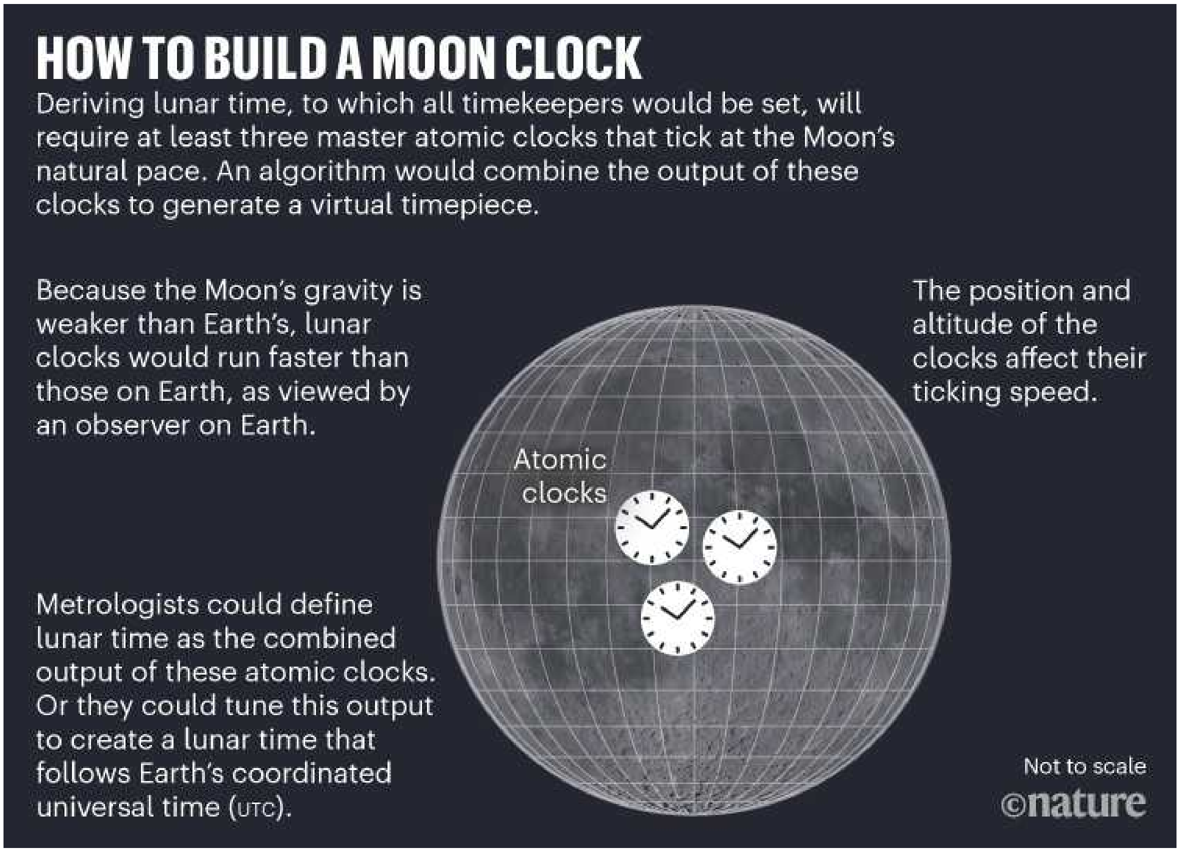
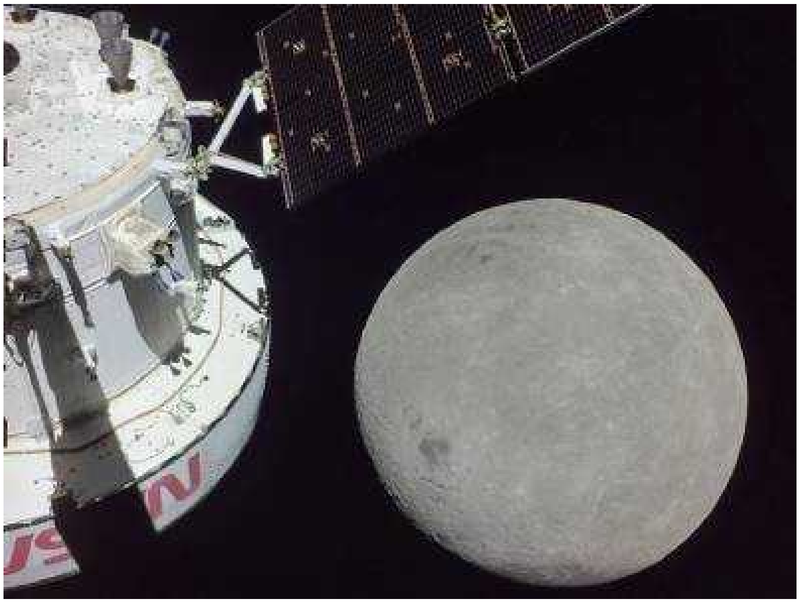


Image credit: NASA

The alternative would be to use the synthesized output of the lunar atomic clocks as the Moon's own independent, continuous time, and to track its relationship to utc. That way, even if the connection with Earth is lost, clocks on the Moon will still agree with each other and allow safe navigation and communications, says Gramling. Establishing an independent time is a model that will also work for the more-distant planets that space agencies are ultimately targeting, such as Mars. Transmitting utc there would be more complicated than to the Moon, she adds.

In this scenario, days on the Moon could even be defined differently from those on Earth, to account for the time from solar noon to solar noon taking an average of 29.5 Earth days. Earth days will always matter to astronauts, given the human need for sleep on a roughly 24-hour cycle. But the definition is something metrologists will need to agree on.



NASA’s Orion spacecraft reaches the Moon — in pictures

Metrologists will also need to decide where on the Moon to place the master clocks. As on Earth, the devices’ altitude will affect ticking speed. The clocks could be in lunar orbit or on the surface, says Hahn. “This is what we are discussing right now with our NASA colleagues.”

Space agencies are also considering other necessary standards — such as which maps of the lunar terrain and coordinate systems to use for navigation — through the Interagency Operations Advisory Group, a council of national space agencies and the United Nations International Committee on GNSS. To make various countries’ systems interoperable, reference systems will have to be agreed internationally, says Gramling.

With ESA’s help, NASA is developing a framework called LunaNet, for which it hopes to get international buy-in. LunaNet consists of a set of rules that would enable all lunar satellite navigation, communication and computing systems to form a single network similar to the Internet, regardless of which nation installs them. Setting lunar time is part of a much bigger picture.

“The idea is to produce a Solar System internet,” says Gramling. “And the first part would be at the Moon.”

Nature **614**, 13-14 (2023)

doi: https://doi.org/10.1038/d41586-023-00185-z

Updates & Corrections

- **Correction 02 February 2023**: This article stated that the effect of clocks ticking more slowly in stronger gravitational fields was due to Einstein's special theory of relativity, when it relates to the general theory.

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00185-z

- NEWS
- 24 January 2023

The next worrisome coronavirus variant could come from China — will it get detected?

Genomic surveillance is crucial for tracking the next ‘variant of concern’, but many countries are winding back their monitoring.

- Dyani Lewis



Strict COVID-19 restrictions have been relaxed in Hong Kong and the Chinese mainland, raising fears that worrying new viral variants will

emerge.Credit: Miguel Candela/SOPA Images/LightRocket via Getty

Fears that the massive surge of coronavirus infections in China could immediately spark the emergence of a troubling new variant are unfounded, say researchers. But that could change in the coming months as more people in the country acquire some natural immunity from infection. More widespread immunity could drive the virus SARS-CoV-2 to evolve ways to evade this immune protection. It remains crucial that variants be tracked, yet scientists question how quickly the next variant of concern will be detected as many countries wind down surveillance efforts.

When China abruptly dropped its zero-COVID policy in December, most of its population had little immunity against the dominant Omicron variant in circulation worldwide. Under such circumstances, the emergence of a dangerous new variant is unlikely, says epidemiologist Jodie McVernon at the Doherty Institute in Melbourne, Australia. There should be less selection pressure for immune-evading variants to emerge in a such a population, she says.

Still, China is ramping up efforts to monitor variants circulating in its population, and has announced plans to get 3 hospitals in each of its 31 provinces to genetically sequence virus samples collected from 15 outpatients, 10 people with severe COVID-19, and all people who have died from COVID-19 each week. But experts are divided on whether these plans will be enough to rapidly detect a concerning variant that could cause new waves of infection and death, in part because many other nations have reduced their genomic monitoring.

However, the European Centre for Disease Prevention and Control has called on European countries to set up random testing of travellers from China, and sequence the virus from all positive samples, so that emerging variants can be detected. Other nations, including the United States, Japan and Australia, have also put surveillance measures in place for travellers from China.

Tracking a virus

Countries track variants by constantly sequencing a proportion of known infections and sharing those sequences in public repositories such as GISAID. During the first two years of the pandemic, most public-health agencies carried out targeted sequencing, monitoring people who had been hospitalized with COVID-19 with the aim of detecting new variants that might cause more-severe illness. Viruses collected from immunocompromised people, who can harbour infections for weeks or months, were also sequenced, because prolonged infections can give rise to heavily mutated viruses[1,2].

Most nations also sequenced a representative sample of viruses from across the community, says Vitali Sintchenko, a microbiologist at the University of Sydney in Australia. In a study he co-authored, the researchers concluded that countries should aim to sequence 0.5% of COVID-19 cases and share those data within 21 days of collecting the samples. That would give them a 34% probability of detecting a new lineage before it infects 100 people[3].

The study, which also looked at sequencing efforts in 189 countries up to the end of February 2022, found that during the first two years of the pandemic, 78% of high-income countries sequenced more than 0.5% of their COVID-19 cases, with some, including Denmark, Japan and the United Kingdom, consistently sequencing more than 5% of cases each week. The earlier such data are gathered and shared, the faster scientists can run laboratory tests to look at the new variant's immune evasion, resistance to antiviral drugs and ability to infect cells, says Sintchenko.

But the testing landscape has changed drastically over the past year, says evolutionary virologist Verity Hill at the Yale School of Public Health in New Haven, Connecticut. Broad-scale population-based screening was feasible in countries such as the United Kingdom because researchers could tap into samples collected at community-based PCR testing facilities. But in many countries authorities are no longer offering such services because of the expense and the decrease in demand, says Hill. And people are increasingly opting to self-test, using rapid antigen tests, or not test at all.

That means that detection of new variants is getting harder everywhere, says Sintchenko.

Red flags

Experts look for mutations in the SARS-CoV-2 spike protein, which allows the virus to enter host cells and is the main target of the body's immune responses. A jump in the number of mutations in a new variant is one thing to watch out for, says Hill. "That's a warning flag," she says. The Omicron variant, which first appeared in a sequence from Botswana, had more than 30 mutations in its spike protein.

The World Health Organization (WHO) only designates a new 'variant of concern' if a variant is better at evading existing immune system protections, causes more severe disease or is much more transmissible than currently circulating variants.

Omicron not only contained many mutations, but also rapidly became a dominant variant in the population, suggesting it was spreading faster than, and out-competing, other variants in the community. The WHO designated Omicron a variant of concern within days of South African researchers alerting the international community to the variant's rapid spread. But that came nearly three weeks after the first Omicron sequence was deposited into GISAID.

The Delta variant was designated a variant of concern in May 2021, seven months after the first known sample was collected in India. The first sign that there could be a concerning new variant around was a rapid rise in case numbers, hospitalizations and deaths in India at the start of 2021. "It's connecting case counts and genetics as much as you can," says Hill.

Wait and see

So far, most of the sequences that China has submitted to GISAID since the beginning of December belong to Omicron subvariants already in circulation elsewhere. There are five new lineages — descendants of those subvariants — but these are unlikely to gain a foothold outside China, because of pre-existing immunity.

But the decreased population-wide surveillance outside China makes it more likely that a variant that emerges in China might initially go undetected, says Hill.

Sintchenko says there are also concerns that China is not sharing enough of its sequences. At a 3 January meeting of the WHO's Technical Advisory Group on Virus Evolution, scientists from the Chinese Center for Disease Control and Prevention presented data based on more than 2,000 genomes collected and sequenced since 1 December 2022. But only around one-quarter of that number — 564 sequences — has been uploaded to GISAID's database over the same period.

A COVID-19 researcher in China, who has asked to remain anonymous to avoid undue attention for weighing in on political matters, says that although current surveillance in China is insufficient, China is building its capacity and ramping up the number of sequences it uploads to GISAID each week.

Nature **614**, 15-16 (2023)

doi: https://doi.org/10.1038/d41586-023-00112-2

References

1. Cele, S. *et al. Cell Host Microbe* **30**, 154–162 (2022).

2. Chaguza, C. *et al.* Preprint at medRxiv https://doi.org/10.1101/2022.06.29.22276868 (2022).

3. Brito, A. F. *et al. Nature Commun.* **13**, 7003 (2022).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00112-2

- NEWS
- 24 January 2023

Scientists petition UCLA to reverse ecologist’s suspension

Sanctions on Priyanga Amarasekare have baffled supporters, who think they are retaliation for speaking out against discrimination.

- Jeff Tollefson



The University of California, Los Angeles, suspended ecologist Priyanga Amarasekare without salary or benefits for one year, and will cut her salary by 20% for two more years.Credit: Al Seib/Los Angeles Times via Getty

In April last year, the Ecological Society of America awarded Priyanga Amarasekare one of the highest honours in the field of ecology: the Robert H. MacArthur Award. A little over two months later, the University of California, Los Angeles (UCLA), placed Amarasekare on a one-year suspension without pay or benefits, and forbade her from accessing her laboratory, maintaining her insect colonies, managing her grants or contacting students. Now scientists from around the world, who call Amarasekare a "highly distinguished ecologist", "a committed teacher and outstanding mentor" and a "tireless advocate for under-represented groups", are calling for her reinstatement.



Scientists question Max Planck Society's treatment of women leaders

The precise allegations that led to her suspension are unknown. UCLA has declined to release them, and barred Amarasekare from discussing the matter publicly. But long-standing tensions between Amarasekare and the university are no secret. Originally from Sri Lanka and one of two women of colour who have tenure in UCLA's ecology and evolution department, Amarasekare has previously accused the university of discrimination for repeatedly denying her promotions that were granted to colleagues. Former students and faculty members who are familiar with the situation think that Amarasekare's suspension was retaliation for speaking out.

Some 315 scientists have raised concerns about her suspension in a petition that was delivered to the university on 23 January, arguing that Amarasekare "has long been denied significant advancement within her department, out of keeping with her contributions to the field". Moreover, the sanctions levied against Amarasekare — including the one-year suspension and a 20% salary reduction for an additional two years — represent "the kind of punishment normally applied only to the most egregious wrongdoings", including scientific misconduct and sexual harassment, the petitioners write.

In the absence of compelling evidence to the contrary, the scientists ask that UCLA rescind the disciplinary actions and fully compensate Amarasekare.

Officials at UCLA say that the university "supports freedom of expression and does not condone retaliation of any sort". They declined to discuss the accusations against Amarasekare or the statements of those who support her, saying the university is "bound to respect the privacy of the numerous individuals involved in this matter". Amarasekare also declined to comment.

A confusing decision

Amarasekare's colleagues told *Nature* that she is the rare ecologist whose research spans theoretical, computational and experimental realms. One project in her laboratory that touches on all of these areas focuses on the impact of climate change on insect communities. "She's really several years ahead of everybody else," says Andy Dobson, an ecologist at Princeton University in New Jersey who helped to lead the petition. Dobson has written letters to support Amarasekare's various applications for promotion at UCLA and says he has been baffled by the university's decisions. "She complained, and most of what's happened seems to be a reaction against that," he says.



Legal win for US scientist bolsters others caught in China crackdown

Nature spoke to several former students and faculty members who defended Amarasekare in administrative hearings in September 2021. Although no one knew the specific details of the charges against her, they all thought she was targeted for speaking out against what she saw as discrimination in the department. In particular, they said Amarasekare vented about her own experience at UCLA on a departmental e-mail listserve created to discuss issues of racism and discrimination in the aftermath of the killing of George Floyd, whose death in May 2020 sparked international protests.

“That’s why she got into trouble. She ended up criticizing pretty much the entire department — with good reason,” says Marcel Vaz, an ecologist at Wilkes University in Wilkes-Barre, Pennsylvania, who was a graduate student in the department at the time. He and other students came forward to support her. “We demanded some explanation,” Vaz says, “but we never got any feedback.”

Peter Kareiva, a former UCLA faculty member who spoke on Amarasekare’s behalf during the administrative proceedings, calls her a brilliant scientist as well as a terrific teacher and student mentor. Kareiva witnessed Amarasekare raise uncomfortable issues and challenge internal policies in faculty meetings. He says she might have made mistakes in terms

of “facilitating harmony” among fellow faculty members, but that her goal was always to improve the department.



How a scandal in spider biology upended researchers’ lives

“I am still incredulous by the punishment levied,” says Kareiva, who now serves as president of the Aquarium of the Pacific in Long Beach, California.

It is unclear what happens next, but scientists contacted by *Nature* are concerned about the impact of the suspension on Amarasekare’s current students, the disruption of her federally funded research and the potentially irretrievable loss of time-sensitive experiments that could provide insights into the ecological impacts of climate change.

As the recipient of the MacArthur award, Amarasekare is expected to discuss this research when she delivers her keynote address at the Ecological Society of America’s annual meeting in Portland, Oregon, in August.

Nature **614**, 16-17 (2023)

doi: https://doi.org/10.1038/d41586-023-00207-w

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00207-w

- NEWS
- 24 January 2023

Bird flu outbreak in mink sparks concern about spread in people

A variant of H5N1 influenza that can spread between mammals could pose an increased risk to people and wild animals.

- Saima May Sidik



The new variant of H5N1 flu had been spreading among mink.Credit: Ole Jensen/Getty

An outbreak of avian influenza on a mink farm in Spain provides the strongest evidence so far that the H5N1 strain of flu can spread from one infected mammal to another.

The outbreak of H5N1 flu, reported in *Eurosurveillance* on 19 January[1], occurred on an American mink (*Neovison vison*) farm in Carral in October 2022. Genetic sequencing showed that the animals were infected with a new variant of H5N1, which includes genetic material from a strain found in gulls, as well as a genetic change known to increase the ability of some animal-flu viruses to reproduce in mammals.

The new variant puts bird flu in "uncharted territory", says Wendy Puryear, a virologist at Tufts University in Medford, Massachusetts. Researchers have warned that, unless careful precautions are taken, the disease might eventually spread among people.

Jumping species

Over the past year, H5N1 has shown an increasing ability to jump from birds to mammals. In the United States, infections have been found in about a dozen species, including raccoons, foxes, seals and grizzly bears.



Why unprecedented bird flu outbreaks sweeping the world are concerning scientists

Until this particular outbreak, all mammalian infections could be attributed to direct contact with virus-contaminated material, says Hualan Chen, a virologist at the Harbin Veterinary Research Institute in China. For example, animals that ingest wild-bird droppings, or that prey on infected animals, can develop the disease. But its spread between mammals “implies that this H5N1 virus may pose a higher risk to public health”, Chen says.

During the first week of October 2022, workers on the affected mink farm noticed that the mink’s mortality rate had increased from a baseline of 0.25% per week to 0.77%, prompting tests on the affected animals for the H5N1 and SARS-CoV-2 viruses. The animals tested positive for H5N1. In the following weeks, more animals fell sick, and the disease seemed to spread from ‘hotspots’ of two to four pens, in which all animals became infected and died. Workers were forced to cull all 51,986 mink on the farm. Eleven farm workers came into contact with the infected mink, but all tested negative for H5N1.

“This species could serve as a potential mixing vessel for the interspecies transmission among birds, mammals and human,” the report’s authors write. “It is necessary to strengthen the culture of biosafety and biosecurity in this farming system and promote the implementation of ad hoc surveillance programs for influenza A viruses and other zoonotic pathogens.”

Preventing spread

Measures taken to prevent the new strain from spreading beyond the farm seem to have been “vigorous, comprehensive and successful”, says infectious-disease specialist William Schaffner at Vanderbilt University Medical Center in Nashville, Tennessee.



COVID mink analysis shows mutations are not dangerous — yet

But Puryear thinks that because the variant contains genetic material from gull flu, it’s likely that at least some of its genetic changes arose in gulls before entering the mink farm. This means that a strain containing those mutations is probably still circulating in the bird population. But for human populations, the outlook is still good: if the new strain did start to infect people, health authorities could probably produce a vaccine quickly, and the antiviral drug Tamiflu (oseltamivir) can reduce the severity of the disease.

The potential risk to wild animals is greater. Bird flu has consistently caused high levels of sickness and death among wild birds and mammals over the past year, and how the new variant will affect that trend remains to be seen. “We just simply don’t know,” says Puryear.

Nature **614**, 17 (2023)

doi: https://doi.org/10.1038/d41586-023-00201-2

References

1. Agüero, M. *et al. Eurosurveillance* **28**, 2300001 (2023).

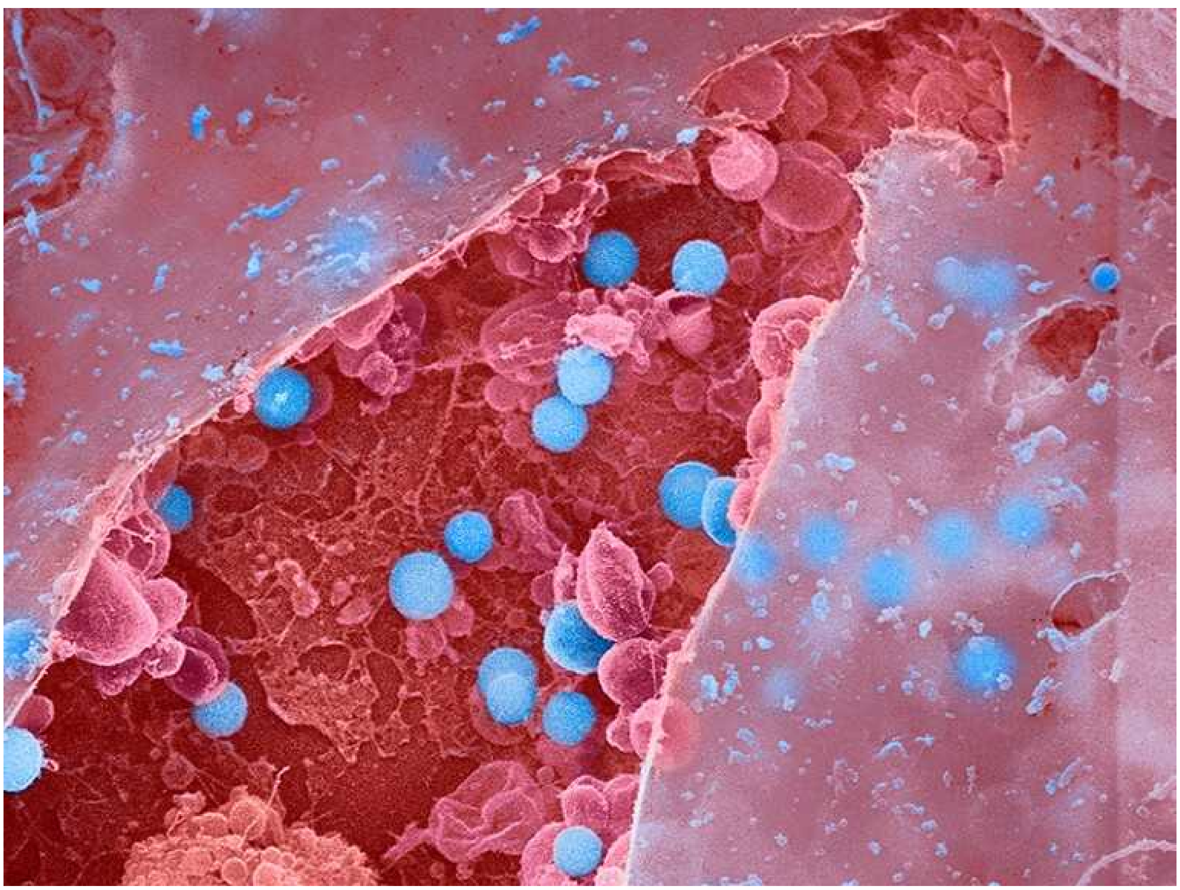
This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00201-2

- NEWS
- 23 January 2023

Massive health-record review links viral illnesses to brain disease

Study ties common viruses such as flu to Alzheimer's and other conditions — but the analysis has limitations, researchers warn.

- Max Kozlov



In this false-colour scanning electron microscope image, influenza virus particles (blue) stand ready to release from a burst epithelial cell

(red).Credit: Lennart Nilsson, Boehringer Ingelheim International GMBH, TT/SPL

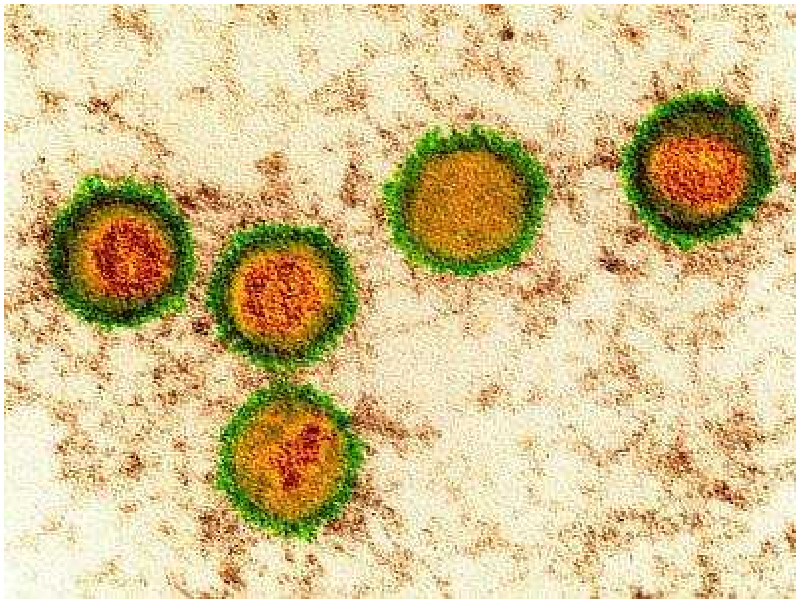
An analysis of about 450,000 electronic health records has found a link between infections with influenza and other common viruses and an elevated risk of having a neurodegenerative condition such as Alzheimer's or Parkinson's disease later in life. But researchers caution that the data show only a possible connection, and that it's still unclear how or whether the infections trigger disease onset.

The analysis, published in *Neuron* on 19 January[1], found at least 22 links between viral infections and neurodegenerative diseases. Some of the viral exposures were associated with an increased risk of brain disease up to 15 years after infection.

"It's startling how widespread these associations seem to be, both for the number of viruses and number of neurodegenerative diseases involved," says Matthew Miller, a viral immunologist at McMaster University in Hamilton, Canada.

Mining health records

This isn't the first time viruses have been linked to neurodegenerative disease. Infection with a type of herpes virus has been associated with the development of Alzheimer's[2], for instance. And a landmark study published in *Science*[3] last year found the strongest evidence yet that Epstein–Barr virus is tied to multiple sclerosis. But many of these past studies examined only a single virus and a specific brain disease.



[The quest to prevent MS — and understand other post-viral diseases]

To understand whether viruses are linked to brain diseases more broadly, Kristin Levine, a biomedical data scientist at the US National Institutes of Health's Center for Alzheimer's Related Dementias in Bethesda, Maryland, and her colleagues analysed hundreds of thousands of medical records to look for instances in which a person had both a viral infection and a brain disease on file.

First, the team examined records from about 35,000 people with brain diseases and about 310,000 people without, sourced from FinnGen, a large Finnish database that includes health information. The team found 45 significant links between infections and brain diseases, and then tested those against more than 100,000 records from another database, the UK Biobank. After this analysis, they were left with 22 significant pairings.

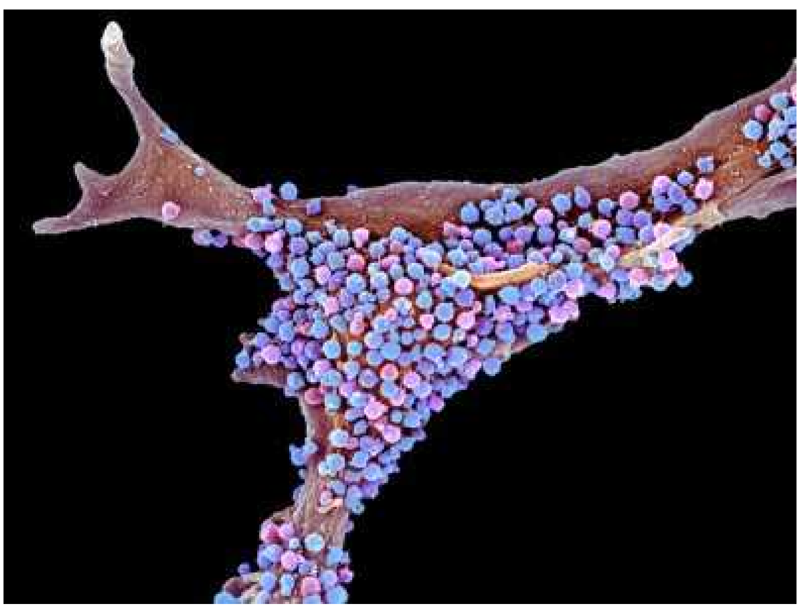
One of the strongest associations was between viral encephalitis, a rare inflammation of the brain that can be caused by multiple types of virus, and Alzheimer's. People with encephalitis were about 31 times more likely to develop Alzheimer's later in life than were people who did not have encephalitis. Most other associations were more modest: people who had a bout of flu that led to pneumonia were four times more likely to develop Alzheimer's than were people who didn't develop the flu with pneumonia.

There were no pairings that suggested a protective link between viral infection and brain disease.

"I'm very excited they're expanding this research broader than what other studies have looked at," says Kristen Funk, a neuroimmunologist at the University of North Carolina at Charlotte who studies the link between herpesviruses and Alzheimer's.

Data shortcomings

Kjetil Bjornevik, an epidemiologist at the Harvard T.H. Chan School of Public Health in Boston, Massachusetts, and an author of the Epstein–Barr paper in *Science*, applauds Levine and her colleagues for bringing more attention to the role of viral infections in brain diseases. But he warns that their approach of using medical records "could be problematic" because they analysed only infections that were severe enough to warrant a trip to a health practitioner. Taking milder infections into account might weaken the associations, he says.

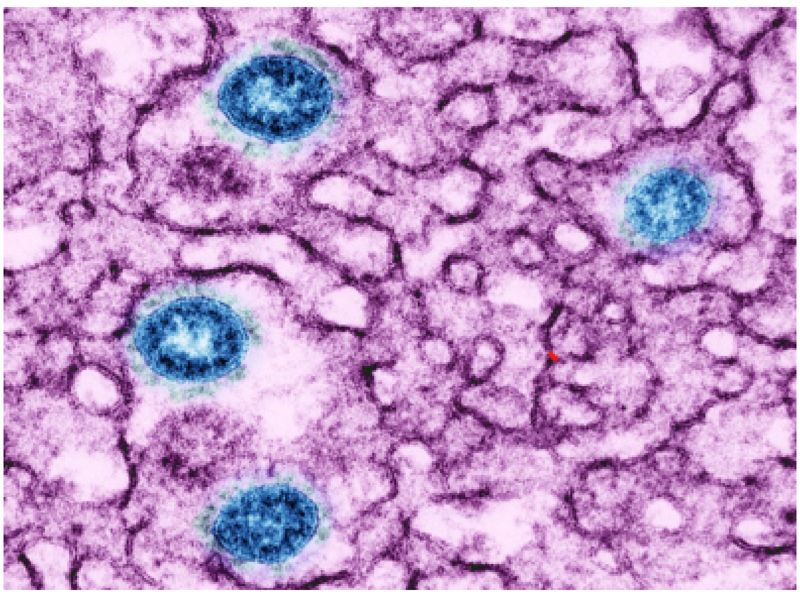


Are infections seeding some cases of Alzheimer's disease?

The data are also sourced almost exclusively from people of European ancestry, which means that the findings might not be applicable to the larger

global population, Funk says. Furthermore, she adds, outside Europe, “certain viruses are more prevalent”, such as Zika or West Nile virus, so the analysis might have missed links between those pathogens and brain disease. Levine acknowledges the limitations of the analysis; the team worked with the data that were available, she says.

These limitations also underscore the difficulty of untangling whether a viral infection leads to neurodegenerative disease, or whether the disease makes a person more susceptible to infection, Bjornevik says. To make it even more tricky, the authors found that the more time that elapsed between the infection and the diagnosis of brain disease, the weaker the link was. The body is known to begin changing years before symptoms of brain disease develop and a diagnosis is made[4], so it’s tough to determine which is causing which, he adds. Another plausible theory is that these viral infections might be accelerating molecular changes in the body that were already ongoing, says Cornelia van Duijn, a genetic epidemiologist at the University of Oxford, UK.



Could long COVID be linked to herpes viruses? Early data offer a hint

If future studies add more weight to the connection between viral infection and brain disease, it could offer health officials a tangible way to delay the onset of neurodegenerative disease. Vaccines exist for many of these viruses, van Duijn says. Because multiple types of dementia are diagnosed late in life

— close to the average life expectancy — if clinicians could postpone disease onset by even a couple of years, that could mean that many people might never develop the disease, she adds.

“It’s not very clear that the infections are causing brain disease,” she says. But viral infections aren’t pleasant, and if there’s any link to brain disease, “I think we owe it to people to prevent them.”

Nature **614**, 18-19 (2023)

doi: https://doi.org/10.1038/d41586-023-00181-3

References

1. Levine, K. S. *et al.* Neuron https://doi.org/10.1016/j.neuron.2022.12.029 (2023).

2. Jamieson, G. A., Maitland, J. J., Wilcock, G. K., Craske, J. & Itzhaki, R. F. *J. Med. Virol.* **33**, 224–227 (1991).

3. Bjornevik, K. *et al. Science* **375**, 296–301 (2022).

4. Swaddiwudhipong, N. *et al. J. Alz. Assoc*. https://doi.org/10.1002/alz.12802 (2022).

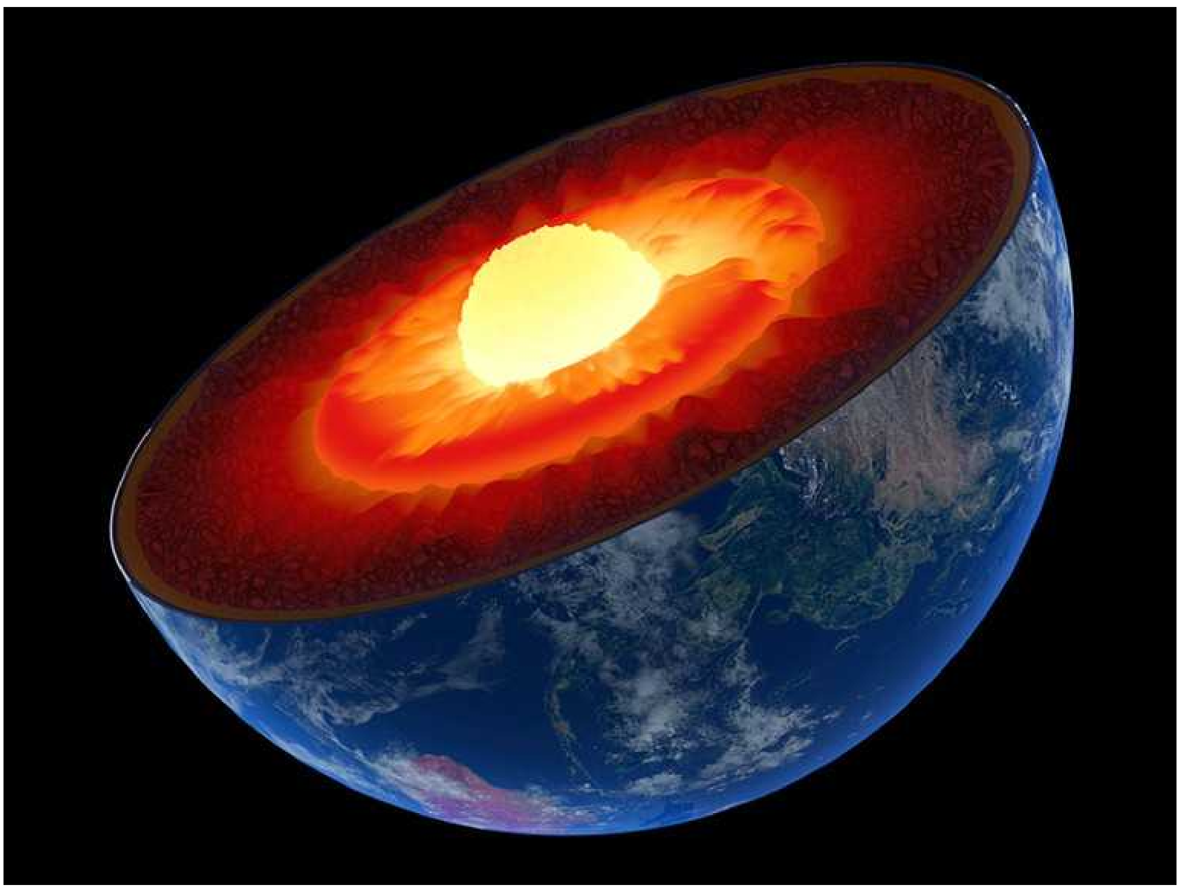
This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00181-3

- NEWS
- 23 January 2023

Has Earth’s inner core stopped its strange spin?

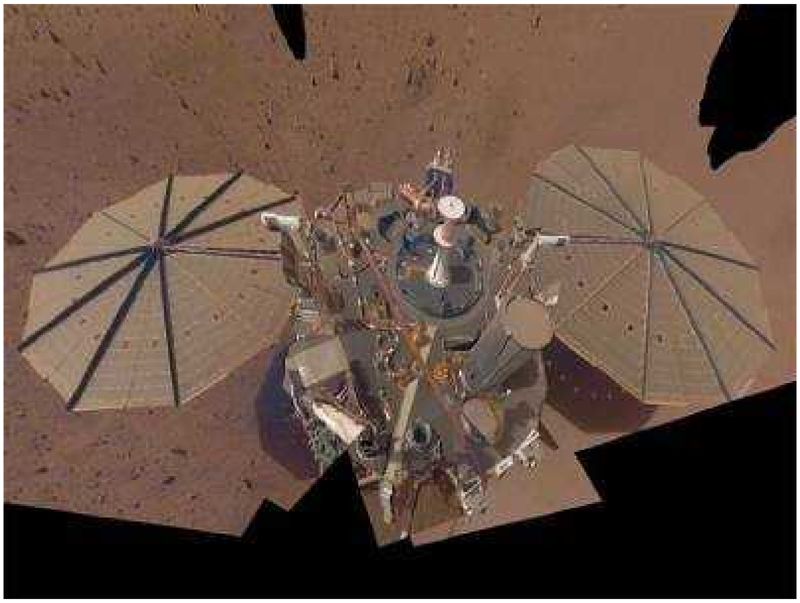
Earthquake data hint that the inner core stopped rotating faster than the rest of the planet in 2009, but not all researchers agree.

- Alexandra Witze



Earth’s inner core is made mostly of solid iron, and can rotate separately from the outer parts of the planet.Credit: Johan Swanepoel/SPL

Thousands of kilometres beneath your feet, Earth's interior might be doing something very weird. Many scientists think that the inner core spins faster than the rest of the planet — but sometime in the past decade, according to a study, it apparently stopped doing so.



Mars's core has been measured — and it's surprisingly large

"We were quite surprised," say Yi Yang and Xiaodong Song, seismologists at Peking University in Beijing who reported the findings today in *Nature Geoscience*[1].

The results could help to shine light on the many mysteries of the deep Earth, including what part the inner core plays in maintaining the planet's magnetic field and in affecting the speed of the whole planet's rotation — and thus the length of a day. But they are just the latest instalment in a long-running effort to explain the inner core's unusual rotation, and might not be the final word on the matter.

"I keep thinking we're on the verge of figuring this out," says John Vidale, a seismologist at the University of Southern California in Los Angeles. "But I'm not sure."

Mysteries of the deep

Researchers discovered the inner core in 1936, after studying how seismic waves from earthquakes travel through the planet. Changes in the speed of the waves revealed that the planet's core, which is about 7,000 kilometres wide, consists of a solid centre, made mostly of iron, inside a shell of liquid iron and other elements. As iron from the outer core crystallizes on the surface of the inner core, it changes the density of the outer liquid, driving churning motions that maintain Earth's magnetic field.



Researchers have learnt about the inner core's rotation by studying earthquakes that originated in the same region, such as the Kuril Islands (shown here), over decades.Credit: Getty

The liquid outer core essentially decouples the 2,400-kilometre-wide inner core from the rest of the planet, so the inner core can spin at its own pace. In 1996, Song and another researcher reported[2] studying earthquakes that originated in the same region over three decades, and whose energy was detected by the same monitoring station thousands of kilometres away. Since

the 1960s, the scientists said, the travel time of seismic waves emanating from those earthquakes had changed, indicating that the inner core rotates faster than the planet's mantle, the layer just beyond the outer core.

Later studies refined estimates of the rate of that 'super-rotation', to conclude that the inner core rotates faster than the mantle by about one-tenth of a degree per year. But not everyone agrees. Other work has suggested that super-rotation happens mostly in distinct periods, such as in the early 2000s, rather than being a continuous, steady phenomenon[3]. Some scientists even argue that super-rotation does not exist, and that the differences in earthquake travel times are instead caused by physical changes on the surface of the inner core[4].

Last June, Vidale and Wei Wang, an Earth scientist also at the University of Southern California, threw another spanner into the works. Using data on seismic waves generated by US nuclear test blasts in 1969 and 1971, they reported that between those years, Earth's inner core had 'subrotated', or rotated more slowly than the mantle[5]. Only after 1971, they say, did it speed up and begin to super-rotate.

A rotational shift

Now, Yang and Song say that the inner core has halted its spin relative to the mantle. They studied earthquakes mostly from between 1995 and 2021, and found that the inner core's super-rotation had stopped around 2009. They observed the change at various points around the globe, which the researchers say confirms it is a true planet-wide phenomenon related to core rotation, and not just a local change on the inner core's surface.

Earth's magnetic field is acting up and geologists don't know why

The data hint that the inner core might even be in the process of shifting back towards subrotation. If so, something is probably happening to the magnetic and gravitational forces that drive the inner core's rotation. Such changes might link the inner core to broader geophysical phenomena such as increases or decreases in the length of a day on Earth.

Still, many questions remain, such as how to reconcile the slow pace of the changes that Yang and Song report with some of the faster changes reported by others. The only way out of the morass is to wait for more earthquakes to happen. A "long history of continuous recording of seismic data is critical for monitoring the motion of the heart of the planet", say Yang and Song.

"We just have to wait," Vidale adds.

Nature **614**, 19 (2023)

doi: https://doi.org/10.1038/d41586-023-00167-1

References

1. Yang, Y. & Song, X. *Nature Geosci*. https://doi.org/10.1038/s41561-022-01112-z (2023).
2. Song, X. & Richards, P. G. *Nature* **382**, 221–224 (1996).
3. Pang, G. & Koper, K. D. *Earth Planet. Sci. Lett.* **584**, 117504 (2022).
4. Yao, J., Tian, D., Sun, L. & Wen, L. *J. Geophys. Res. Solid Earth* **124**, 6720–6736 (2019).
5. Wang, W. & Vidale, J. E. *Sci. Adv.* **8**, eabm9916 (2022).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00167-1

- NEWS EXPLAINER
- 27 January 2023

Will a new wave of RSV vaccines stop the dangerous virus?

Pfizer, GSK and Moderna are ahead in the race to produce vaccines against respiratory syncytial virus — what will be their impact?

- Liam Drew



Respiratory syncytial virus infects the lungs where it can cause serious illness in very young and very old people.Credit: NIAID/NIH/SPL

Last week, the pharmaceutical company Moderna announced promising results from a trial of people over 60 for its vaccine against a potentially fatal disease: respiratory syncytial virus (RSV). The pharmaceutical companies Pfizer and GSK announced similar phase III results for this age group last year — both are now awaiting regulatory approval. Pfizer has also filed for approval to give its vaccine during pregnancy, following results showing that it protected newborn babies against RSV. Assuming they will be approved soon, what is likely to be the impact of these first RSV vaccines?

RSV usually causes mild cold-like symptoms, but can be dangerous in very young and very old people. Cases have soared in recent months, putting stress on health-care systems. In babies aged between around one month and one year, RSV is globally the second-highest cause of death, behind malaria. For adults, the data are sparser: a study[1] in 2022 estimated that around 6,500 people in the United States die of RSV annually, with the highest rates in people over 65.

But the products being developed for each age group present different opportunities and challenges. "The RSV vaccine field is not one field but two," says Louis Bont, a paediatrician specializing in infectious disease at University Medical Centre Utrecht in the Netherlands. "One is to protect babies and one is to protect the elderly."

The quest for an RSV vaccine began in the 1960s, and there have been notable failures along the way. A 2013 breakthrough[2] in deciphering the structure of the protein that the virus uses to anchor to and enter human cells paved the way for the current wave of RSV vaccines.

Stable protein

Moderna's mRNA-based vaccine stimulates the production of a stabilized version of this protein, whereas the Pfizer and GSK vaccines inject synthetic versions of it directly. In trials, their efficacies have been remarkably similar, although Bont expects differences to emerge in their initial protection they offer and the durability of immunity after they are rolled out on a larger scale.

RSV vaccines for people over 60 are the closest to being available. But if they are approved, it’s not clear what their uptake among this age group will be.

Although initial COVID-19 vaccination rates were high in older people, Christine Shaw, vice-president and portfolio head for respiratory vaccines at Moderna in Cambridge, Massachusetts, says “that was a pandemic emergency”. For RSV, “the awareness and the fear, it's not the same”. Uptake of COVID-19 boosters has fallen off in older populations and influenza jabs are often not as popular as public-health experts would like. Shaw says Moderna is exploring whether it can combine its RSV vaccine with its COVID-19 booster and/or a flu shot to improve convenience.

It is also not clear how strongly countries will recommend the vaccine for over-60s. In South Africa, for example, the impact of an RSV vaccine for older people is likely to be relatively low, says Cheryl Cohen, an epidemiologist and respiratory-disease specialist at the University of the Witwatersrand in Johannesburg, South Africa. Older people make up a small proportion of its total population and flu vaccines have not achieved good uptake in this group.

Benefits for babies

When it comes to vaccines for infants, Cohen says it’s a different story: “It’s one of the top new products that we’re looking for.” The benefits she lists — preventing a large number of hospitalizations and deaths, relieving stress on hospitals and the fact that pregnant people are in frequent contact with health-care systems, meaning vaccines can be integrated into routine care — apply across low-, middle- and high-income countries.

“I think that in ten years, we’re going to talk about the RSV vaccine in moms, and we’re going to look back and say, ‘Look at what this moment in public-health history was’,” says Alejandra Gurtman, vice-president for vaccine clinical research and development at Pfizer in New York City.

Vaccinating during pregnancy generates high levels of maternal antibodies, which transfer through the placenta into the fetus, protecting babies through

their first few months, explains Gurtman. Pfizer's trial, which it reported on last November, showed that babies born to people vaccinated during pregnancy were substantially less likely to develop RSV infections that required clinical intervention. Moderna says it is planning similar trials, but GSK halted its pregnancy trial last year owing to safety concerns.

The optimism about RSV prevention in children isn't just about vaccinating pregnant people. An alternative is prophylactically injecting newborn infants with antibodies against RSV proteins: one such antibody, developed by AstraZeneca and Sanofi, was approved for widespread use in November 2022. As well as providing choice for parents, Bont says, the antibodies could be important for premature babies who did not receive sufficient antibodies from their vaccinated parents.

The Bill & Melinda Gates Foundation and Gavi, the Vaccine Alliance, are working to improve access to RSV vaccines and monoclonal antibodies in low- and middle-income countries. But Bont and Cohen stress that continued advocacy and political prioritization will be needed to ensure equitable access.

Nature **614**, 20 (2023)

doi: https://doi.org/10.1038/d41586-023-00212-z

References

1. Hansen, C. L., Chaves, S. S., Demont, C. & Viboud, C. *JAMA Netw. Open.* **5**, e220527 (2022).

2. McLellan, J. S. *et al. Science* **340**, 1113–1117 (2013).

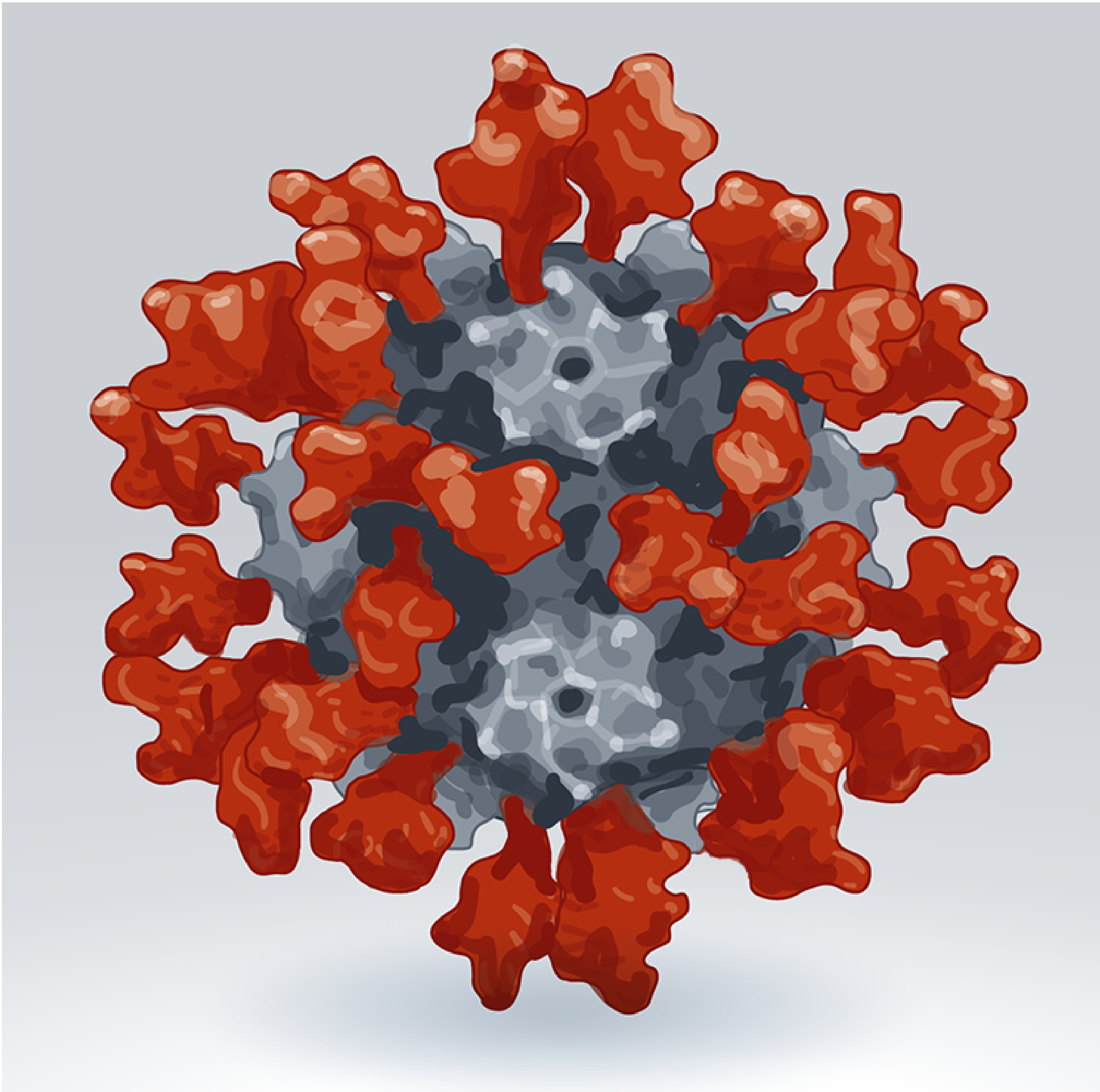
This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00212-z

- NEWS FEATURE
- 01 February 2023

The next generation of coronavirus vaccines: a graphical guide

New technologies might provide more potent or broader immunity — but will have to fight for market share.

- Ewen Callaway



A nanoparticle vaccine from South Korean firm SK bioscience, approved in 2022.Nik Spencer/*Nature*. Adapted from A. C. Walls *et al. Cell* **183**, 1367–1382 (2020).

Vaccines against the coronavirus SARS-CoV-2 have been given to billions of people to protect them from COVID-19, and have saved more than 20 million lives. But viral variants can evade some of the immunity provided by the original vaccines. As a result, vaccine developers around the world are working on dozens of ‘next-generation’ COVID-19 vaccines: not just updates of the first versions, but ones that use new technologies and platforms.

These vaccines are a diverse group, but the overarching aim is to deliver long-lasting protection that is resilient to viral change. Some could protect against broader classes of coronavirus, including ones that have yet to emerge. Others might provide more potent immunity, might do so at lower doses, or might be better at preventing infection or transmission of the virus.

Here’s what to expect of this next generation of vaccines.

Why do we need more vaccines?

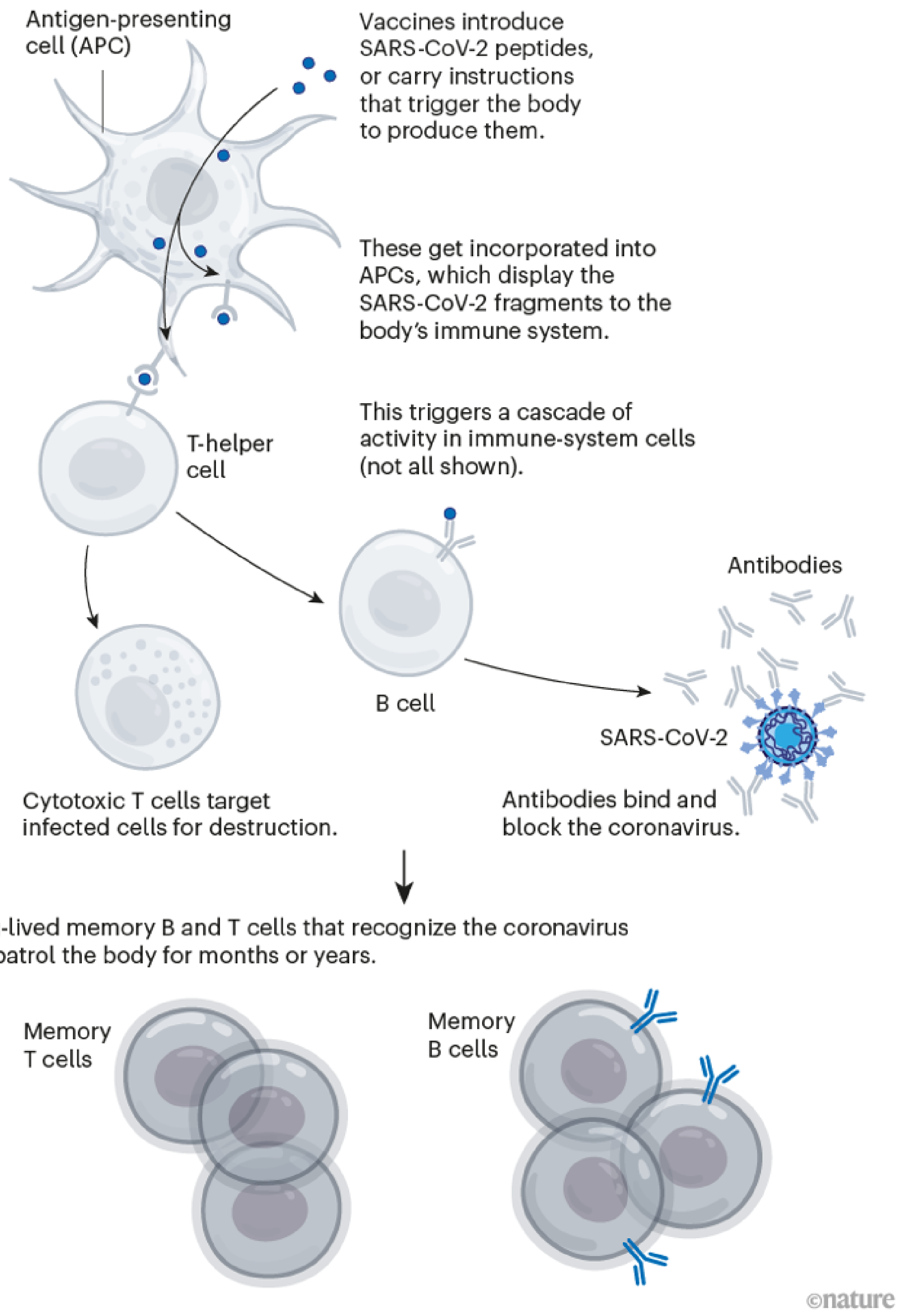
In a word: evolution. The first approved COVID-19 vaccines were tested for protection against versions of SARS-CoV-2 that had not changed much since the virus was first identified. These vaccines come in different types — some are composed of messenger RNA, others are inactivated versions of the coronavirus itself or some of its proteins — but all work by exposing the body to antigens (portions of the virus) to provoke an immune response without causing disease.

Broadly speaking, this immune response comes from B cells, which produce antibodies that can block SARS-CoV-2 from infecting cells, and from T cells, which can destroy infected cells (and support other immune responses).

The vaccinations also generate a pool of ‘memory cells’ for prolonged immunity, even after initial antibody levels dwindle. On subsequent infection, memory B cells begin proliferating and differentiating into cells that churn out more antibodies (see ‘How coronavirus vaccines protect against SARS-CoV-2’).

HOW CORONAVIRUS VACCINES PROTECT AGAINST SARS-COV-2

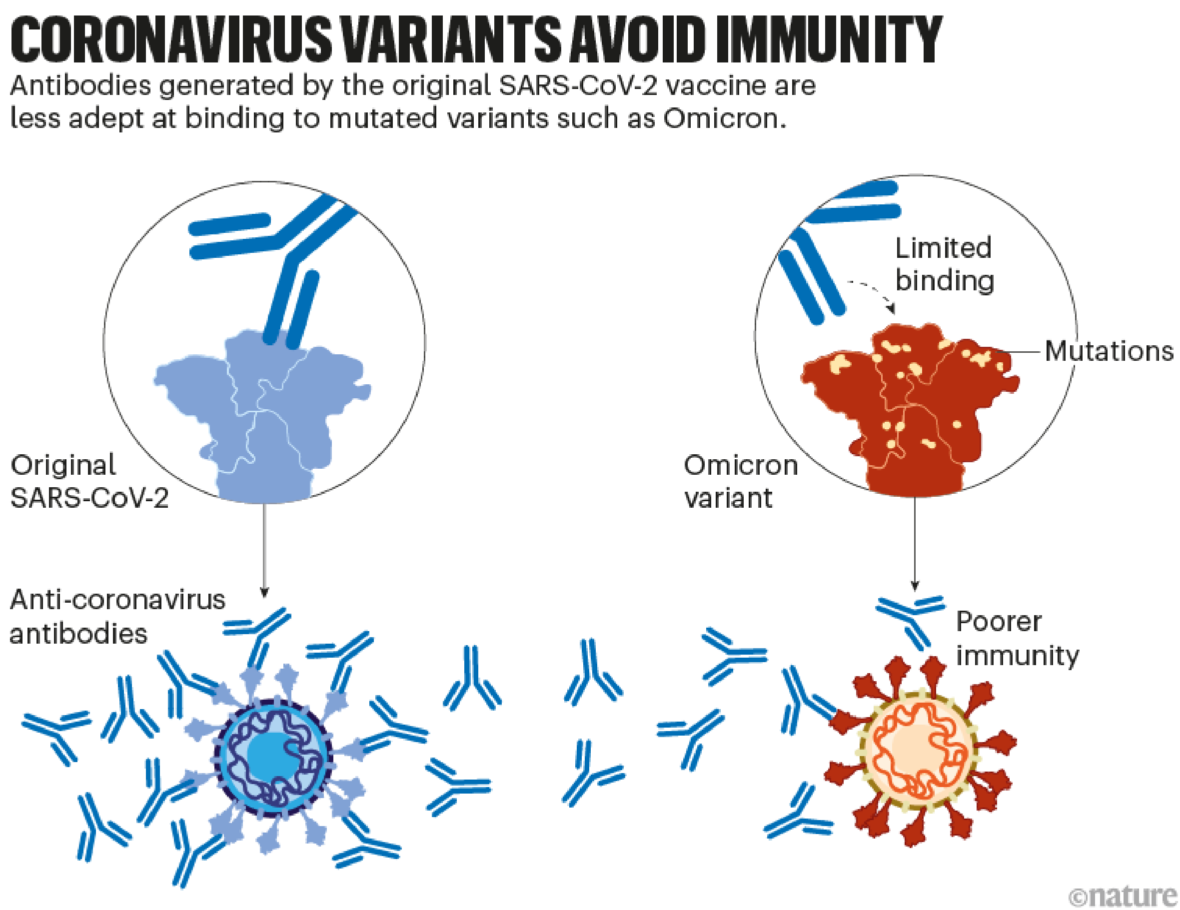
COVID-19 vaccines expose our bodies to portions of the SARS-CoV-2 virus, provoking an immune response (see simplified process below), with B cells and T cells readied to fight infection.



©nature

Nik Spencer/*Nature*.

Although these vaccines provide long-lasting protection against severe disease, the protection they offer against viral infection dwindles in months. And variants of SARS-CoV-2, such as Omicron, have since evolved with mutations that allow them to escape some of this immunity. For instance, memory responses generated by the initial vaccines produce antibodies that don't latch on to Omicron as easily. That contributes to the reduced protection against infection (see 'Coronavirus variants avoid immunity').



Nik Spencer/*Nature*.

A second generation of vaccines has already been introduced to boost immunity against the Omicron variant. It's likely that [further, variant-specific updates to vaccines] will follow, to try to keep up with viral evolution — although it's not clear whether the protection they offer will be

particularly long-lasting as immunity wanes and SARS-CoV-2 evolves further.

As a result, research teams are taking several approaches to develop new vaccines.

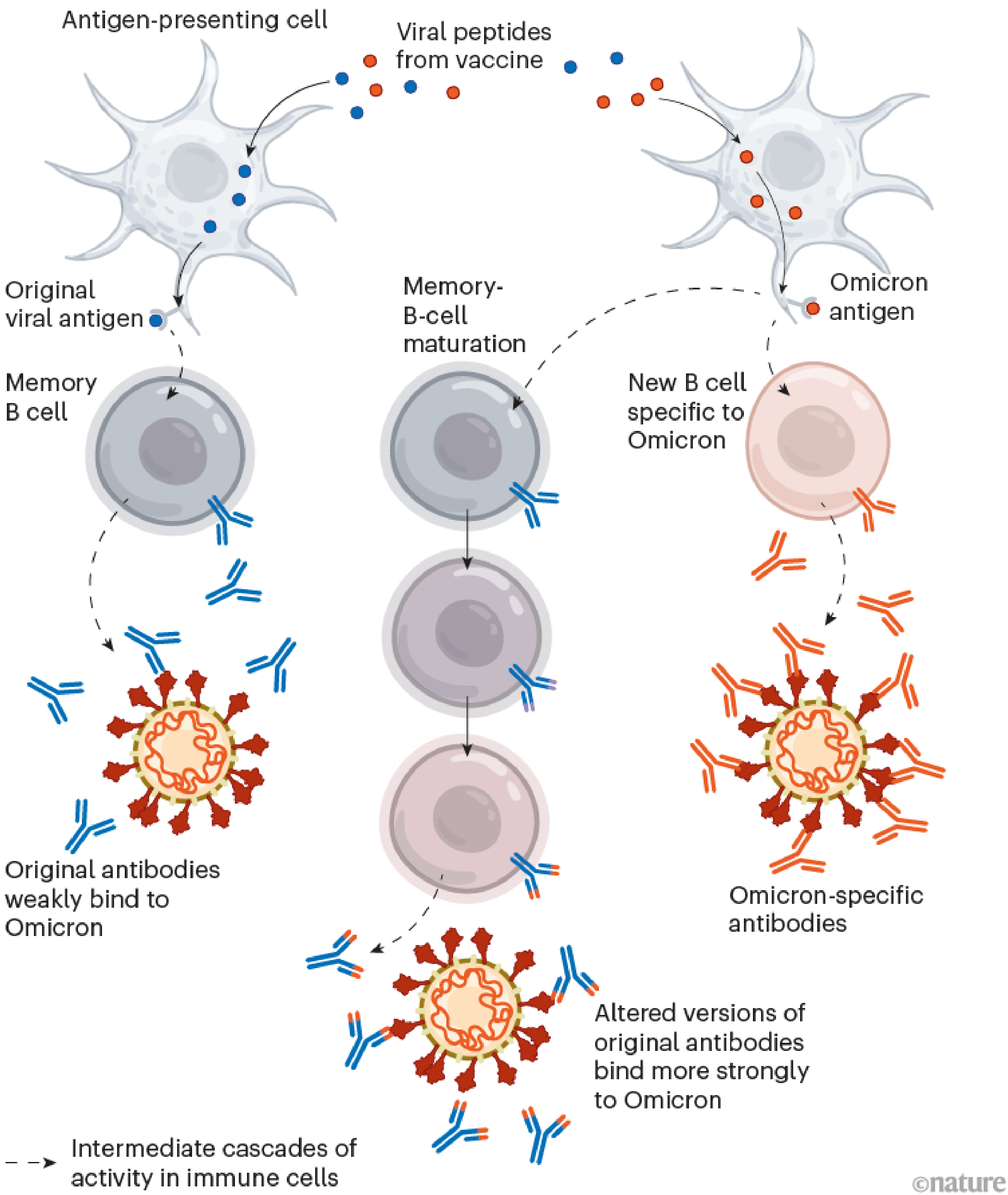
Updated vaccines

To tackle SARS-CoV-2 variants, the vaccine developers Pfizer–BioNTech and Moderna introduced updated mRNA vaccines last year. These are called bivalent, because they encode molecules of the spike protein from the original virus and from Omicron. (The spike protein is what SARS-CoV-2 uses to bind to cells.)

The bivalent vaccines work in several ways. Like other COVID-19 booster shots, they stimulate the memory B cells already established by previous vaccines; some of this cellular response leads to antibodies that can recognize Omicron. Their potency can strengthen over time, too: when presented with Omicron's spike, memory B cells go through an evolutionary 'training' process of mutation and selection, producing a pool of B cells that encode antibodies that bind more tightly to Omicron's spike. Finally, the Omicron components of bivalent vaccines also recruit new B cells that produce their own antibodies (see 'Updated vaccines').

UPDATED VACCINES

'Bivalent' vaccines present portions of both the original virus and its variants (such as Omicron), to try to keep up with the evolving virus. In response to a booster dose of this vaccine, some existing memory B cells spark off the production of more antibodies; others mature, leading to cells able to make antibodies that bind to Omicron more strongly; and some new B cells make Omicron-specific antibodies.



©nature

Nik Spencer/*Nature*.

These effects might mean that a bivalent booster provides better protection against Omicron than does a booster dose of the original vaccine. But it's still unclear how substantial that advantage is in practice.

Some developers, including Pfizer–BioNTech, are also working on combination vaccines to protect people against COVID-19 and other diseases — most commonly influenza. Nearly all are in the early stages of development.

Broadly protective vaccines

Updates to COVID-19 vaccines will always be a step or two behind the evolving virus. Scientists hope to develop 'broadly protective' vaccines that can target future SARS-CoV-2 variants — and even related coronaviruses.

The goal of some of these vaccines is to generate an immune response against particular regions of the spike protein that are conserved across SARS-CoV-2 variants and some related coronavirus species, meaning that they tend not to mutate in new variants. One region of interest is the receptor-binding domain (RBD), which binds to the ACE2 receptor protein on human cells and is targeted by some of the body's most potent infection-blocking antibodies.

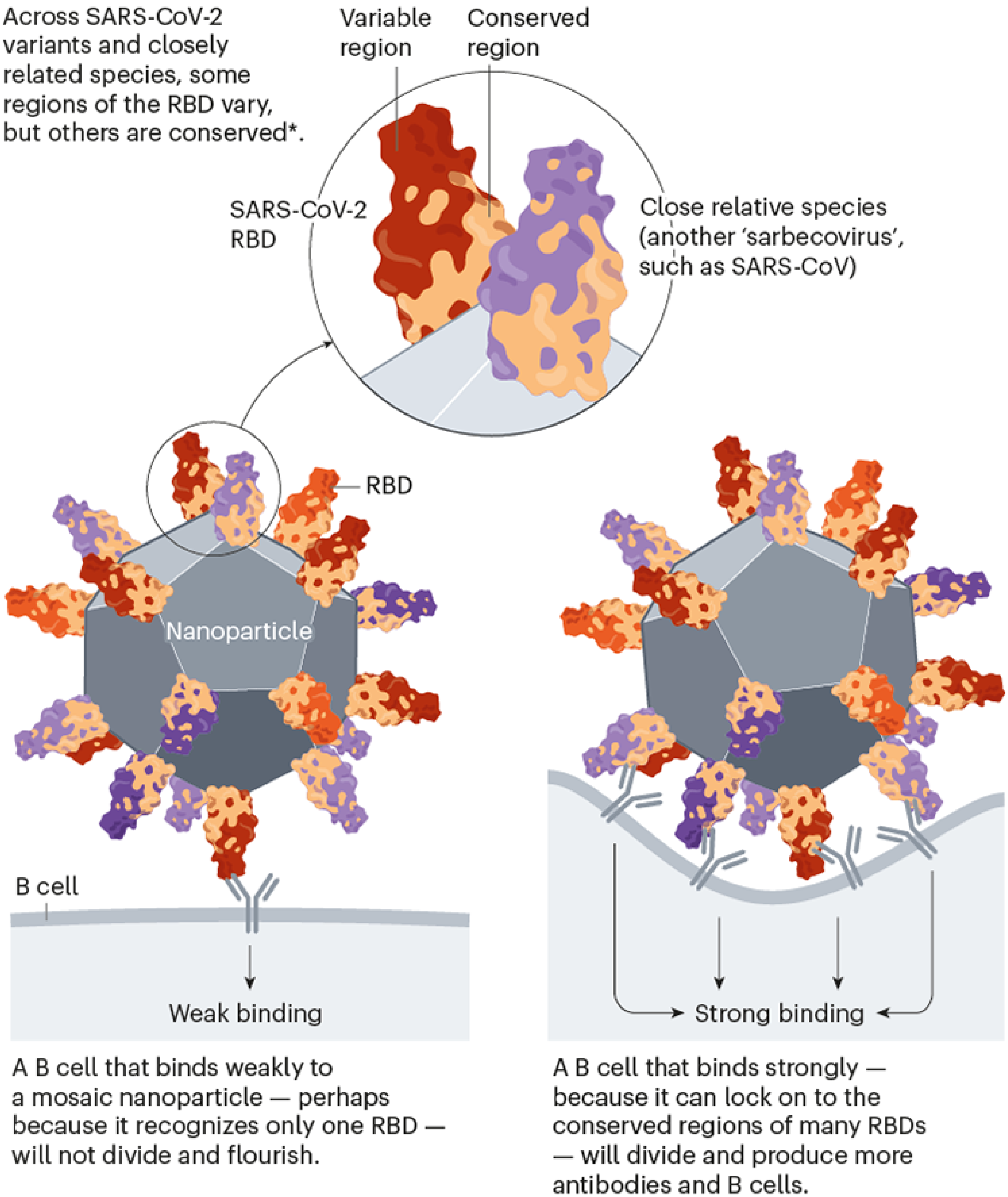
At least two teams, at the University of Washington in Seattle and at the California Institute of Technology (Caltech) in Pasadena, are making 'mosaic' vaccines: nanoparticles dotted with RBDs from SARS-CoV-2 and coronaviruses from the same family (called sarbecoviruses), such as SARS-CoV and others isolated from bats.

When a B cell recognizes more than one RBD on these mosaic nanoparticles — latching on to conserved regions from multiple virus species — it binds strongly. This, in turn, triggers that B cell to multiply and produce more antibodies (as well as memory B cells to fight future infections). B cells that recognize an RBD from just one viral species bind weakly, and do not generate this response. Researchers hope that using mosaic nanoparticles

will result in an enriched pool of antibodies that can recognize multiple RBDs across coronavirus species (see ‘Broader protection?’).

BROADER PROTECTION?

Some vaccines might be able to target many future SARS-CoV-2 variants. 'Mosaic' vaccines stud nanoparticles with receptor-binding domains (RBDs) from multiple coronaviruses. They might be able to produce antibodies that can bind to — and so protect against — many related coronaviruses.



*At the resolution shown here, the RBDs for different sarbecoviruses look the same.

©nature

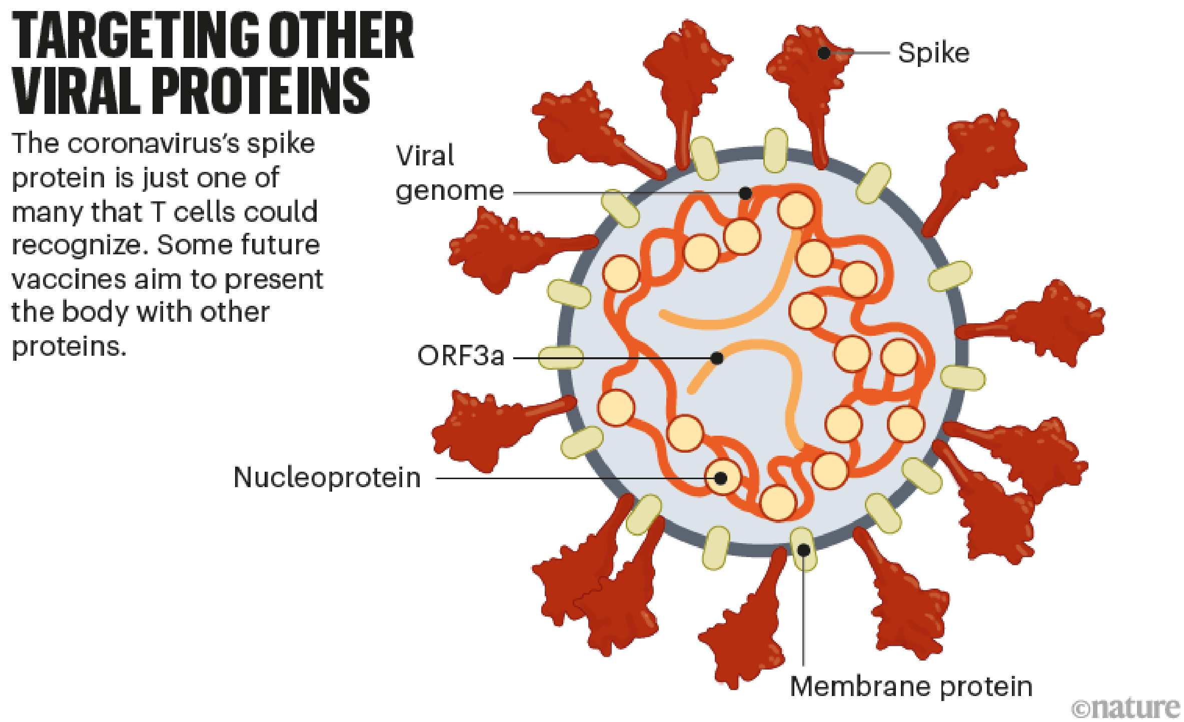
Nik Spencer/*Nature*. Adapted from Caltech (https://go.nature.com/3R6GWJ3)

Animal studies suggest that these vaccines do trigger protective responses against diverse sarbecoviruses (see, for example, A. A. Cohen *et al. Science* 377, eabq0839; 2022). The first clinical trials are set to begin in the next two years.

Going beyond spike

Many first-generation COVID-19 vaccines prompt an immune response only against SARS-CoV-2's spike protein.

But some next-generation vaccines deliver other viral proteins as well, in the hope of generating a more diverse immune response that safely mimics the protection conferred by infection. This approach could also mitigate the impact of new spike variants (see 'Targeting other viral proteins').



Nik Spencer/*Nature*. Adapted from Stanford Univ. (https://go.nature.com/3WV2FB6)

The spike protein is the main target of antibody-making B cells. But T cells that destroy infected cells can recognize many other SARS-CoV-2 proteins. For this reason, vaccines that deliver other proteins could help protect people whose immune systems do not generate strong antibody responses. Such vaccines might also be more resilient to viral evolution, because non-spike proteins tend to vary less between variants.

The US biotechnology company Gritstone is developing one such vaccine: it delivers instructions for several SARS-CoV-2 proteins using mRNA vaccine technology. Meanwhile, Texas biotech company Vaxxinity is developing a protein-based vaccine that would expose the body to multiple antigens. The company says it plans to apply for UK and Australian authorization this year, after a phase III trial showed the vaccine was safe and prompted a strong antibody response when used as a booster.

New platform designs

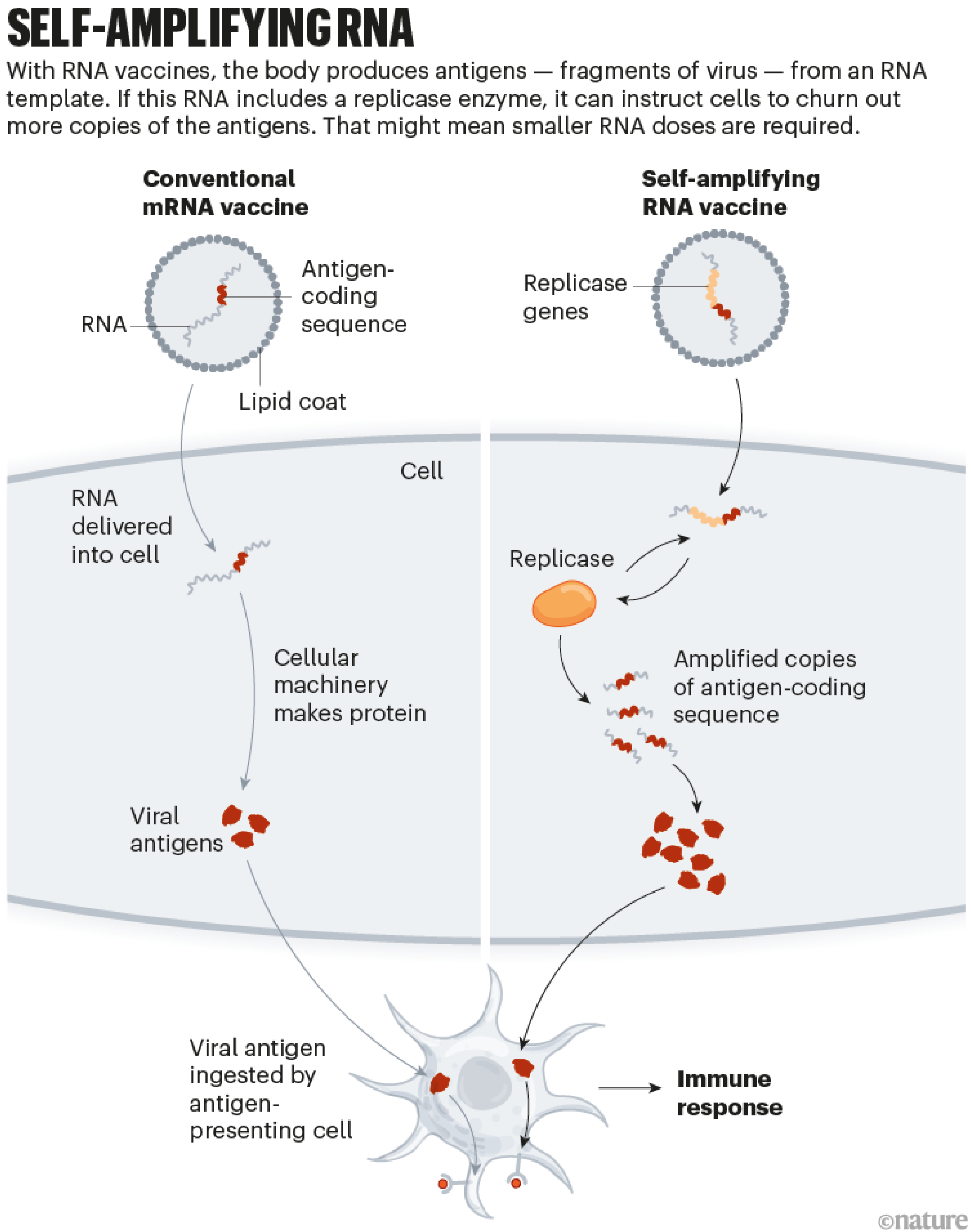
Another way of categorizing next-generation vaccines is by the method of delivery into the body. Existing vaccines use one of at least four approaches: nucleic-acid vaccines (mostly mRNA) instruct cells to make the SARS-CoV-2 spike protein; inactivated vaccines use versions of the coronavirus itself; protein vaccines are composed of the spike protein or its RBD; and viral-vector vaccines use modified viruses to shuttle instructions for the spike protein into cells. Next-generation vaccines could involve tweaks to these designs or changes to delivery mechanisms that might improve performance.

Self-amplifying RNAs

mRNA vaccines helped to turn the tide of the pandemic, particularly in wealthy countries, where the vast majority of doses have been sold. A twist on this technology might make vaccines cheaper and even more potent, while minimizing side effects.

The vaccines developed by Pfizer–BioNTech and by Moderna (with the US National Institute of Allergy and Infectious Diseases) consist of mRNA instructions for a modified version of spike packaged in a fatty nanoparticle. In an updated version of this technology, self-amplifying RNA (saRNA)

vaccines also include instructions for an enzyme that instructs cells to churn out more copies of spike (see ‘Self-amplifying RNA’).



Nik Spencer/*Nature*.

This means that a smaller — and potentially cheaper — dose of saRNA vaccines could achieve the same or even a stronger immune response, compared with conventional mRNA vaccines. A smaller initial dose might also reduce side effects.

One saRNA vaccine, developed by US firm Arcturus Therapeutics, completed a phase III trial in April 2022; the company has now started another phase III trial in Japan that, it says, could lead to an application for authorization there. Gritstone is using saRNA technology to deliver additional SARS-CoV-2 proteins in a candidate T-cell vaccine that has completed a phase I trial.

Proteins on nanoparticles

Several protein-based COVID-19 vaccines have been authorized globally, including one made by US biotech firm Novavax. Their low cost and ease of production makes them appealing; they are usually made of stabilized forms of the entire SARS-CoV-2 spike protein or its RBD.

A new class of these vaccines is made of proteins that self-assemble into a soccer-ball-shaped structure, studded with spike or RBD. The repetitive arrangement of the viral molecules, mimicking an actual virus, generates an especially potent immune response.

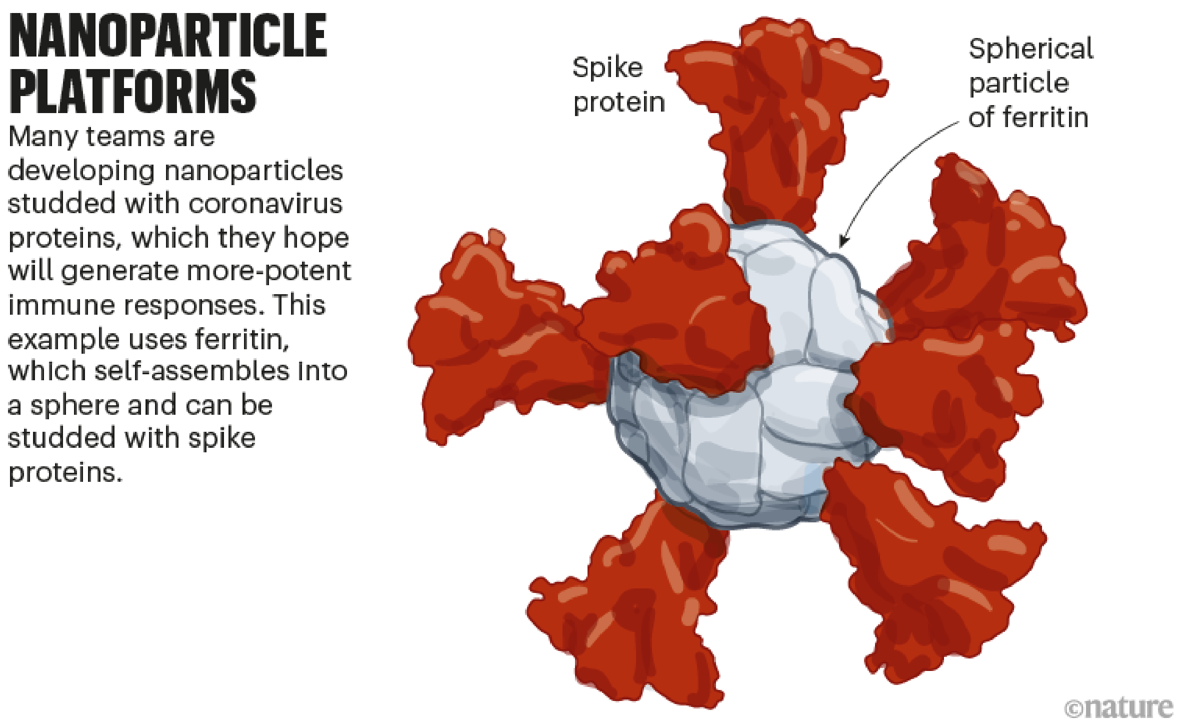
The ‘mosaic’ vaccines developed by Caltech and the University of Washington (which are studded with RBDs from multiple kinds of coronavirus) are one example of this effort.

Another nanoparticle vaccine has already been approved: in April 2022, South Korean regulators authorized a vaccine, also developed at the University of Washington, containing RBDs from the original version of SARS-CoV-2. A phase III trial showed that the vaccine boosted antibody responses to levels that were several times higher than those generated by the viral-vector vaccine developed by AstraZeneca and the University of Oxford, UK, which uses a chimpanzee adenovirus encoding spike antigens.

However, the South Korean company developing the vaccine, SK biosciences, said in late 2022 that it had paused production amid low

demand for the vaccine in South Korea.

A team led by researchers at the US Walter Reed Army Institute of Research in Silver Spring, Maryland, is developing another protein nanoparticle vaccine, using an iron-carrying protein called ferritin. This self-assembles into a spherical particle, and is then studded with the full SARS-CoV-2 spike protein. It is currently being tested in an early-stage trial (see ‘Nanoparticle platforms’).

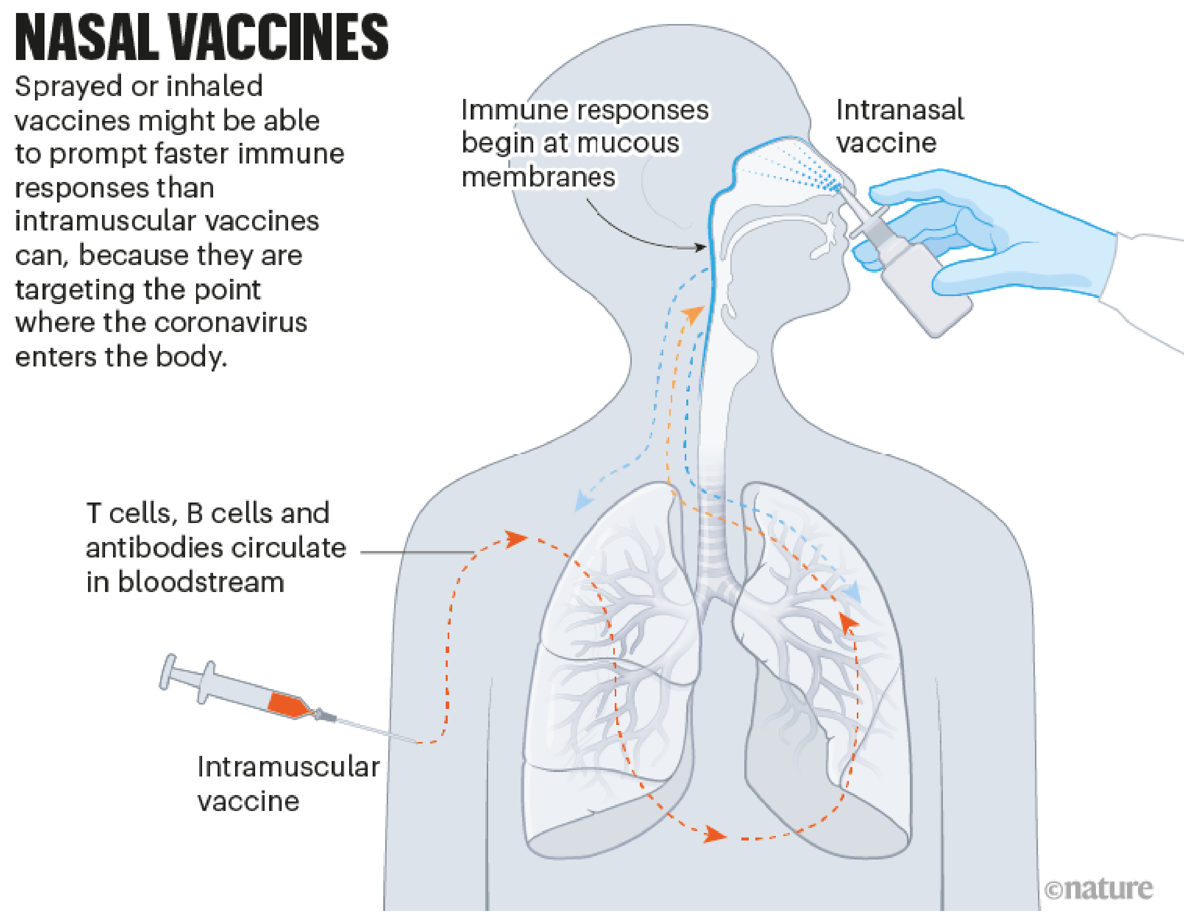


Nik Spencer/*Nature*. Adapted from Stanford Univ. (https://go.nature.com/3WV2FB6) and WRAIR (https://go.nature.com/3JTX3XU)

Nasal vaccines

Some COVID-19 vaccines are inhaled as a mist through the nose or mouth, or as nasal drops. By prompting immune responses at the point where SARS-CoV-2 enters the body — in the thin mucous membranes that line the nose and mouth — these vaccines could, in theory, stop the virus before it spreads.

Data from animal studies suggest this might be possible, and at least five nasal vaccines have already been approved for use — two in China and one each in India, Iran and Russia. But there are no data yet on whether these vaccines are better than injections at cutting down infection or transmission of the virus (see 'Nasal vaccines').

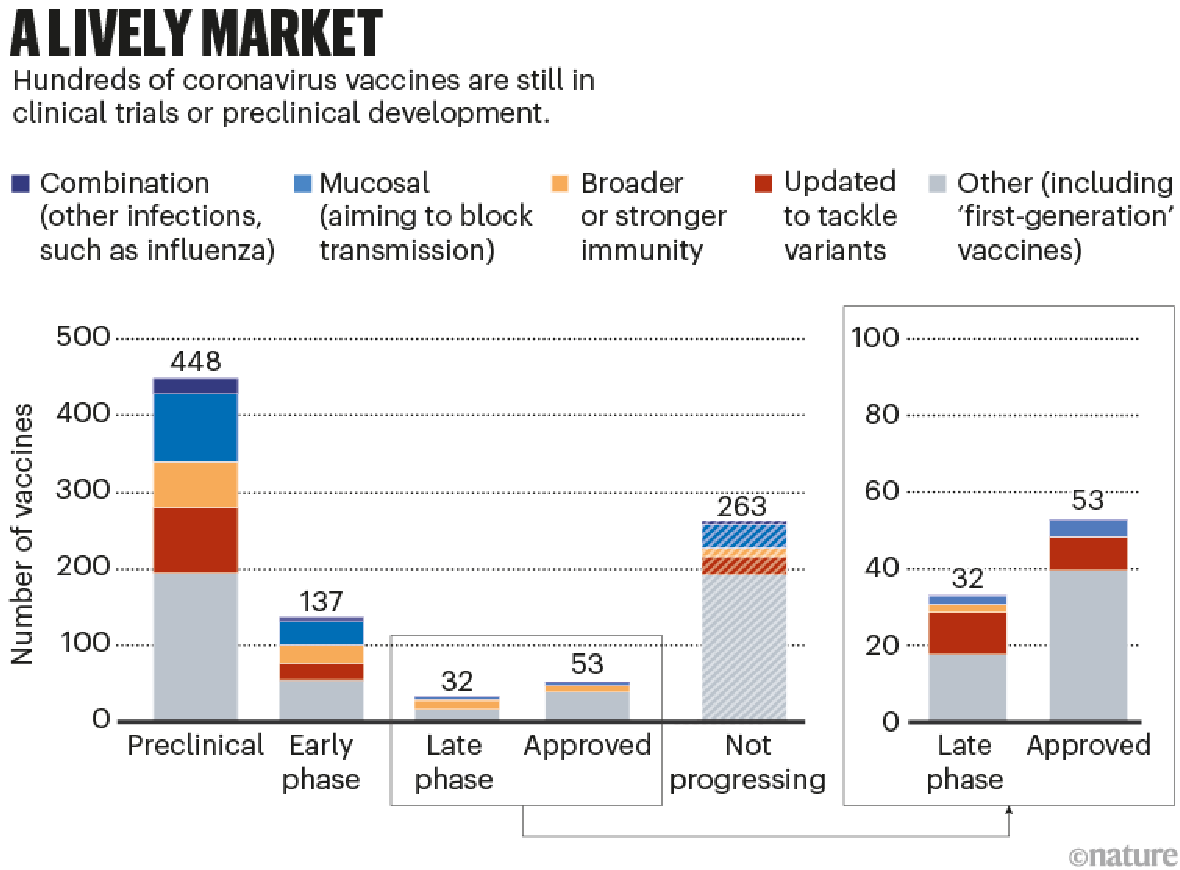


Nik Spencer/*Nature*.

A key challenge to the development of these and other next-generation COVID-19 vaccines is proving that they offer genuine improvements over existing jabs, says Melanie Saville, executive director of vaccine research and development at the Coalition for Epidemic Preparedness Innovations (CEPI), an Oslo-based foundation that is a leading funder of next-generation COVID-19 vaccines.

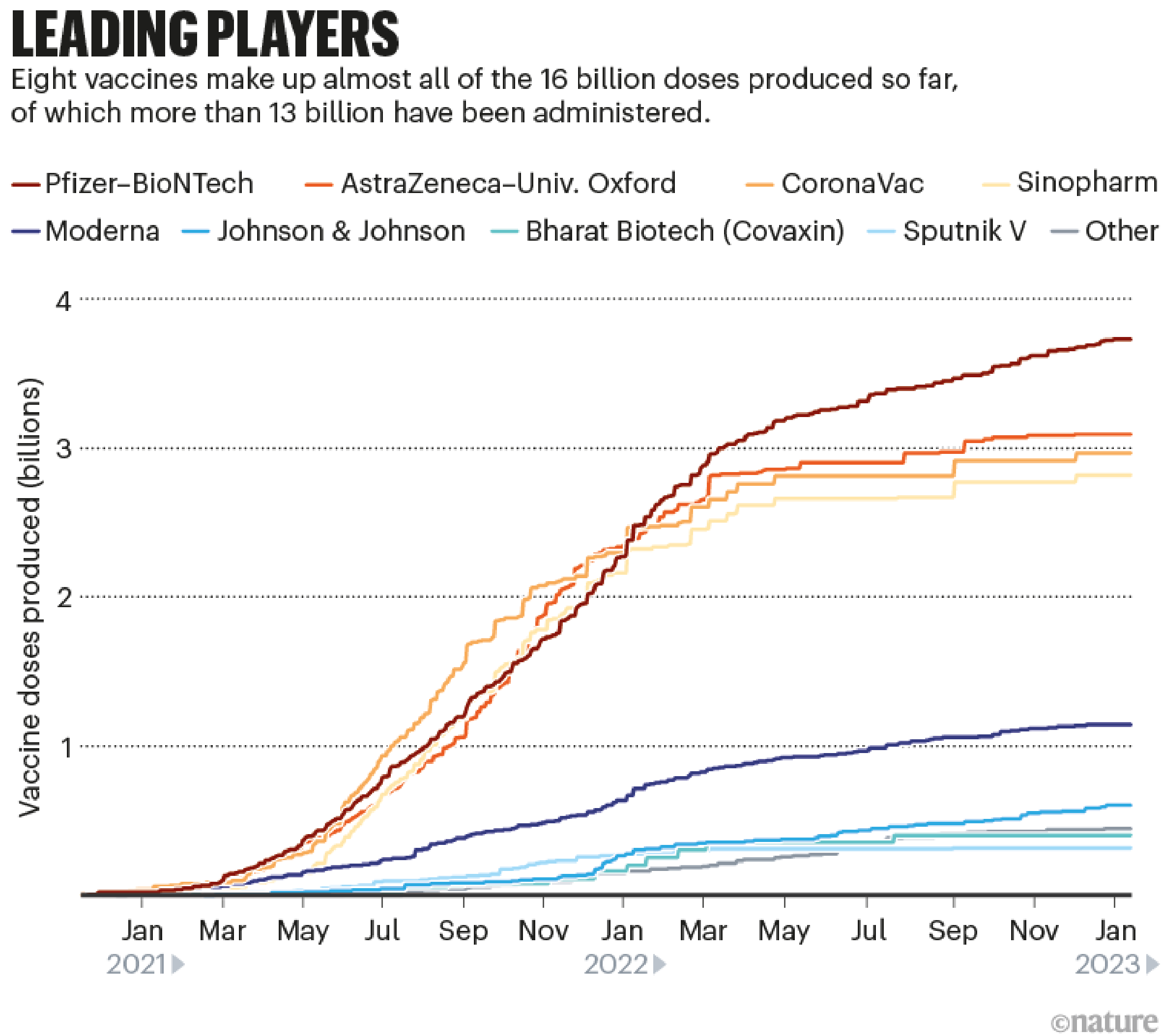
Stiff competition

All the next-generation vaccines will have to fight for market share. More than 50 vaccines have already been approved, and there are hundreds in early- and late-stage clinical trials; hundreds more have been abandoned (see ‘A lively market’).



Nik Spencer/*Nature.* Source: Airfinity

Among the approved vaccines, just a few dominate the doses that have been administered (see ‘Leading players’).



Nik Spencer/*Nature*. Source: Airfinity

Despite the flurry of research, current mRNA jabs such as the Moderna and Pfizer–BioNTech ones are likely to hold sway, says Matt Linley, analytics director at Airfinity, a life-sciences information firm in London. The fast development of bivalent vaccines that included an Omicron component showed that these vaccines could be adapted quickly. If another update is needed, “mRNA vaccinations would be market leaders in being able to react quickly”, says Linley.

COVID-19 is no longer seen as the all-encompassing emergency it once was, and so developers and regulators will move more slowly compared with the breakneck pace of the first-generation vaccines, adds Saville. “We

shouldn’t rush these through, because these need to be the types of vaccine that will be durable in the long term for COVID-19.”

But even if work on new vaccine technologies doesn’t directly pay off against COVID-19, it could still support efforts to combat other diseases, Saville says, such as CEPI’s work on a ‘vaccine library’ for different virus families to improve preparation for future threats.

Nature **614**, 22-25 (2023)

doi: https://doi.org/10.1038/d41586-023-00220-z

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00220-z

Books & Arts

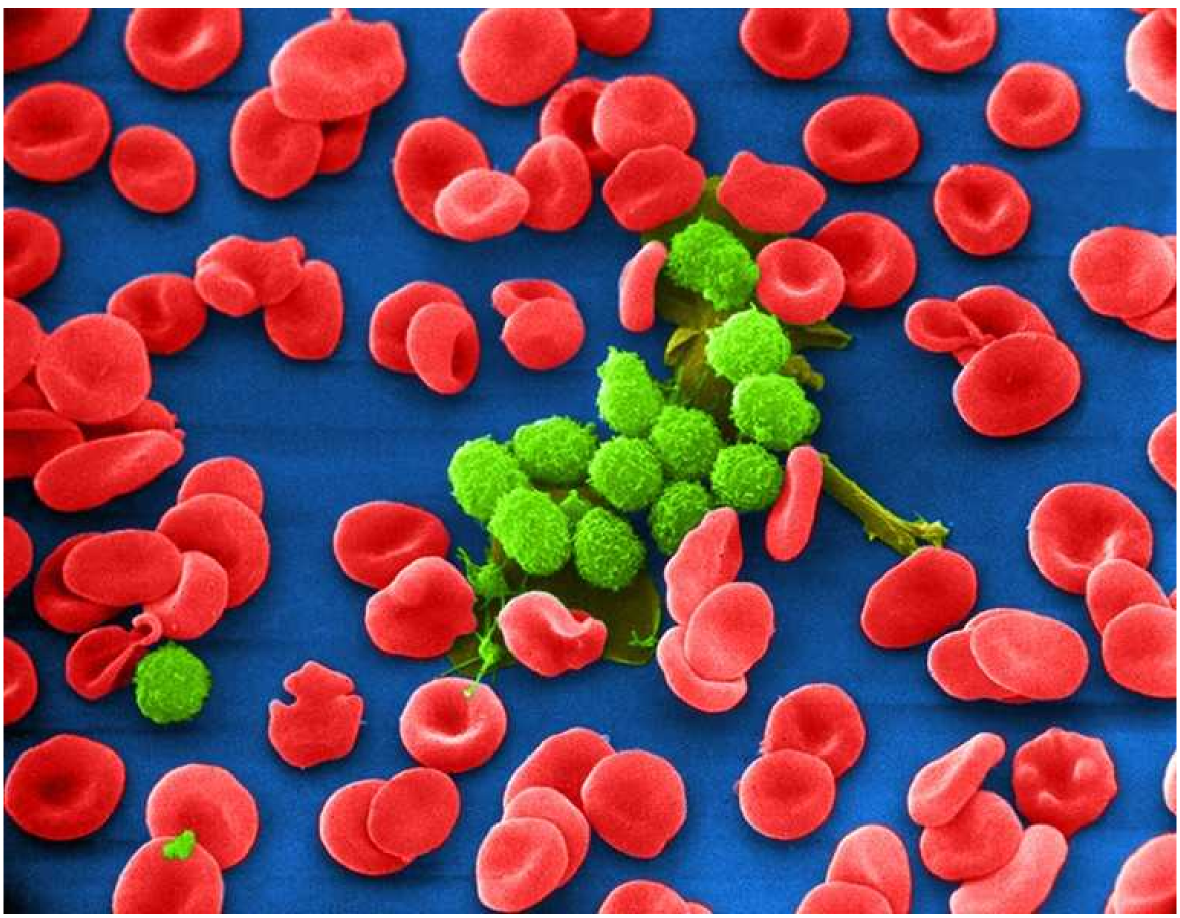
- **A crash course in biotech success — and failure** [30 Jan 2023]
  • The unlikely discovery of a life-changing leukaemia drug uncovers harsh realities of profit and loss.
- **The history of US health inequity, and a vision for the Moon: Books in Brief** [23 Dec 2022]
  • Andrew Robinson reviews five of the best science picks.

- BOOK REVIEW
- 30 January 2023

A crash course in biotech success — and failure

The unlikely discovery of a life-changing leukaemia drug uncovers harsh realities of profit and loss.

- Heidi Ledford

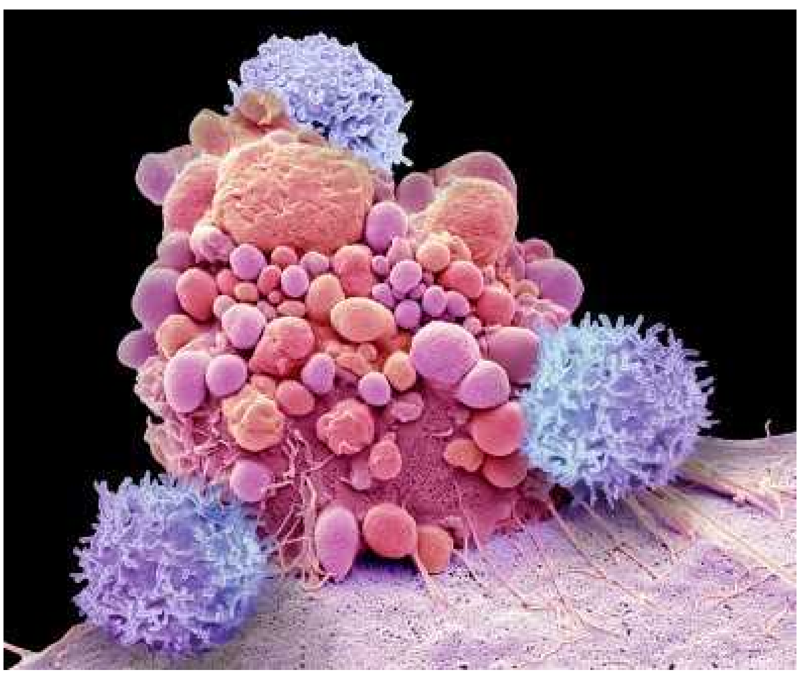


The cancer drug ibrutinib curbs the uncontrolled proliferation of B cells (green).Credit: Keith R. Porter/SPL

For Blood and Money: Billionaires, Biotech, and the Quest for a Blockbuster Drug *Nathan Vardi* W. W. Norton (2023)

Even in the most experienced hands, the vast majority of candidate drugs fail. So it hardly seems a recipe for success when a serial entrepreneur whose previous business highlights include investments in needlepoint kits and bakeries takes over a struggling biotechnology firm.

Improbably, it was. Although Robert Duggan did not achieve his initial goal of developing a treatment for brain cancer, the disease that claimed his son's life, his company went on to produce a transformative leukaemia drug called ibrutinib, and Duggan became a billionaire. Yet, despite the financial happy ending, the story that veteran business reporter Nathan Vardi tells in *For Blood and Money* is no fairy tale. The book is an engaging tutorial on the harsh realities of drug development, including a warts-and-all look at the outsize role of chance and money — and how science often ends up taking a back seat to finance.



Last-resort cancer therapy holds back disease for more than a decade

Duggan took over Pharmacyclics in 2008 when the company, based in Sunnyvale, California, seemed to be on its last legs. It had just weathered a series of bruising clinical-trial results that spelled the end for its most

promising prospective treatment, a drug intended to shrink brain tumours that not only failed to do that, but also sometimes turned trial participants' skin green.

Potential for success

Years earlier, Pharmacyclics had happened to buy a few bargain-priced assets from another company, including a compound that binds to and inhibits a protein called Bruton's tyrosine kinase. The compound was originally meant to be a research tool, not a medicine. In the wake of the failed brain cancer trials, however, Pharmacyclics decided to put the inhibitor into play. It reasoned that, because the kinase was important for immune cells called B cells, blocking the protein could be a way to quench cancers that are caused by uncontrolled B-cell proliferation, such as some forms of lymphoma and leukaemia.

Staff members at Pharmacyclics worked around the clock to test the compound amid a whirl of company politics and pressure from early investors. Word got out of Duggan's eccentric leadership style — heavily influenced by his faith in Scientology, hard-driving personality and experience in consumer-oriented industries with less regulation than for biotech firms such as Pharmacyclics. "Those who had heard of it, through the Silicon Valley biotechnology grapevine, knew it to be a weird place," Vardi writes.

A bitter pill

Even so, much of Pharmacyclics' broader story is a common biotech parable. Academic scientists with an interest in drug development will benefit from the reality check. Science was just one of many factors that determined ibrutinib's fate, among them money, competition, intellectual-property law, regulatory requirements and health-care economics. Researchers often did not pull the strings — in fact, few were still around to reap the rewards when the cash came rolling in. The puppeteers were the investors who were willing to put their money on the line and who had little-to-no formal training in science.



[Why children have to wait years for new drugs]

Vardi spends much of the book detailing how Wayne Rothbaum, a savvy micromanaging stock trader and early investor in Pharmacyclics, influences the company's direction, including key scientific decisions. Tensions rise and Duggan fires a few key executives who had worked around the clock on the drug and has them escorted out of the building. A few weeks later, Vardi says, one borrows a friend's credentials to sneak into a Pharmacyclics presentation at a prominent oncology conference. She sits in the audience, weeping as she learns that the clinical trial she had helped to design was a success.

And a success it was: ibrutinib slowed cancer growth in two-thirds of people with chronic lymphocytic leukaemia, the most common form of cancer in adults, and it did so with fewer toxic side effects than standard chemotherapy. Vardi describes a participant who woke up one morning to find that his cancer-swollen lymph nodes had shrunk and the constant pain they caused him was gone. His first thought was that he must have died. Another participant travels with his full supply of ibrutinib in his pockets, worried that an emergency might separate him from his bags.

Ultimately, Pharmacyclics reached the climax that awaits many successful biotech companies: being bought out by a larger pharmaceutical firm that has the cash and infrastructure needed to mass-produce and sell a drug while

complying with regulatory requirements. In 2015, AbbVie in North Chicago, Illinois, acquired Pharmacyclics for US$21 billion. The hard work of a select few company employees was rewarded with millions of dollars; the investors who took on a financial risk to bankroll the operation got billions. Some of the scientists and executives at Pharmacyclics were bitter over the magnitude of this difference, and Vardi highlights the people who were left out of completely, including the academic researchers who ran the clinical trials and the scientists who first designed ibrutinib.



Could CAR-T-cell therapy offer hope to children with cancer?

But readers outside the biotech bubble might reserve their anger for another aspect of the story. When ibrutinib hit the market in 2013, a course cost $131,000 per year in the United States. Some people need to take the drug for years. Ibrutinib might be less physically toxic than various other cancer treatments, but it is financially toxic for many of the people who need it.

For Blood and Money does an excellent job of highlighting the complexities and expense, both financial and personal, of drug development. But, given that more than 40% of people with cancer in the United States have burnt through their life savings within two years of diagnosis, the giant payouts at the end are perhaps the bitterest pills to swallow.

Nature **614**, 27-28 (2023)

doi: https://doi.org/10.1038/d41586-023-00221-y

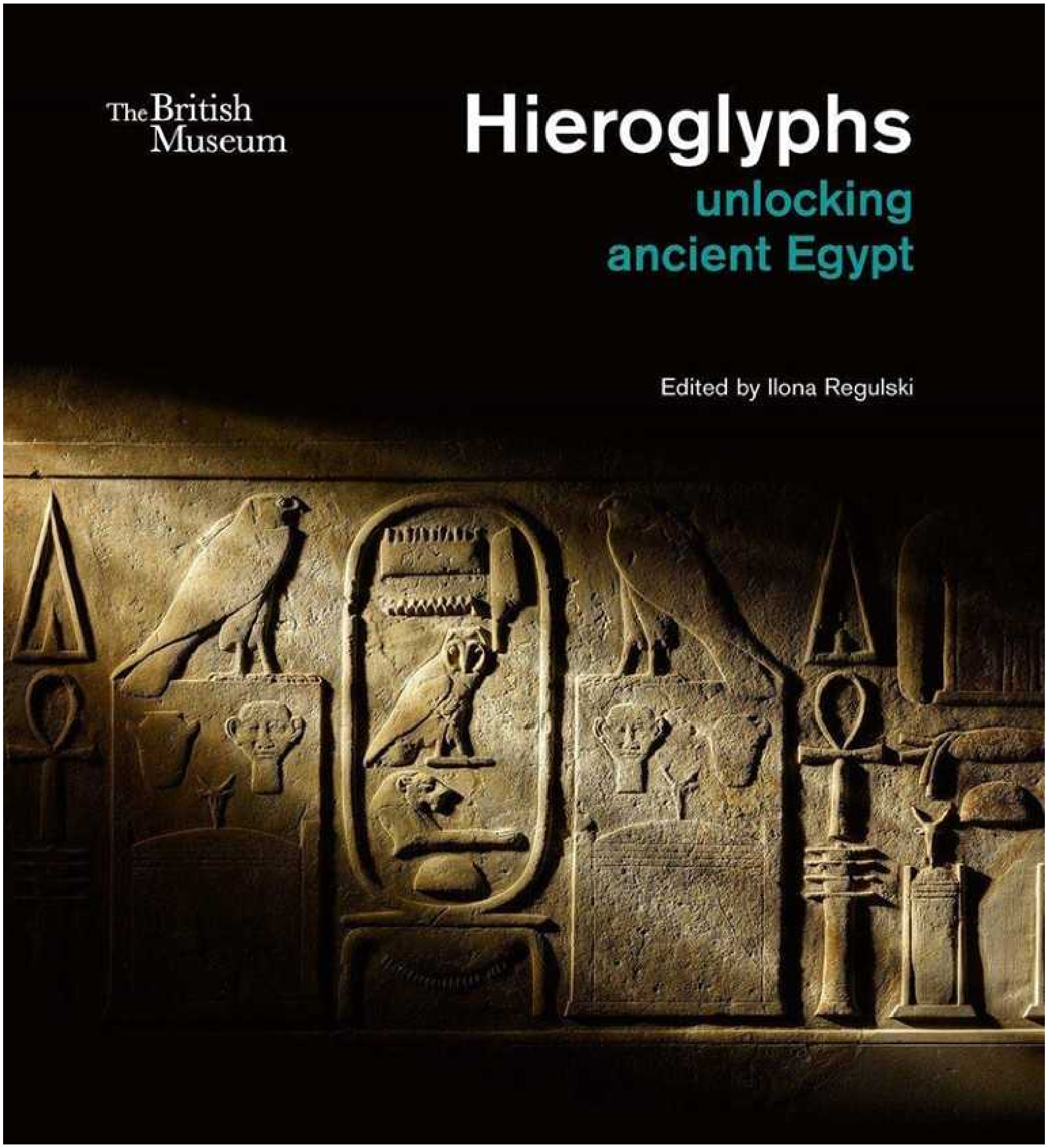
This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00221-y

- BOOK REVIEW
- 23 December 2022

The history of US health inequity, and a vision for the Moon: Books in Brief

Andrew Robinson reviews five of the best science picks.

- Andrew Robinson

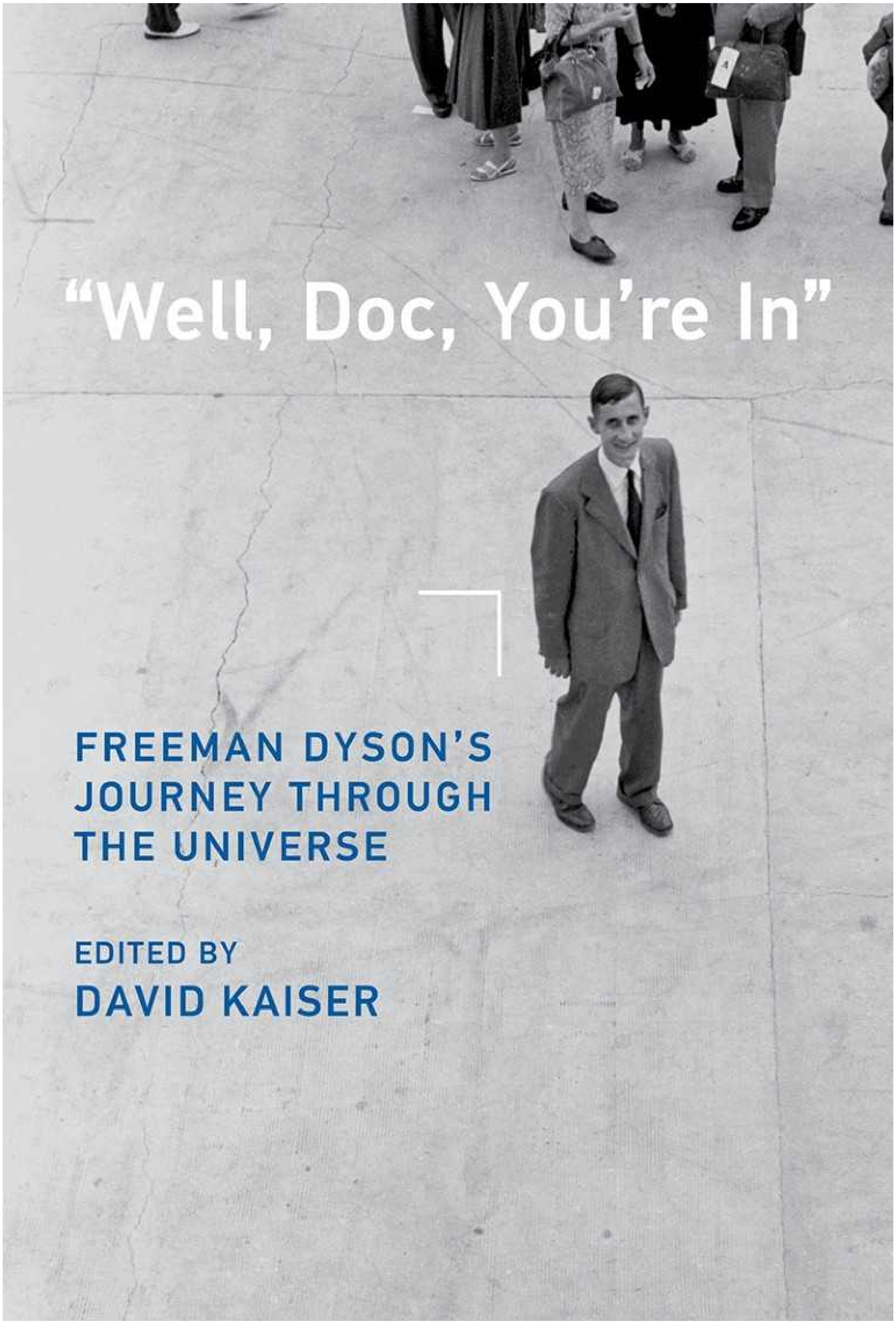


Hieroglyphs

Ed. Ilona Regulski *British Museum Press* (2022)

The 196 bc Rosetta Stone is the most visited object in the British Museum, London — a source of fascination for its role in explicating Egyptian hieroglyphics. Its decipherment, initiated by polymath Thomas Young and achieved by philologist Jean-François Champollion starting in 1822, “changed our understanding of the ancient world”, says Egyptologist Ilona

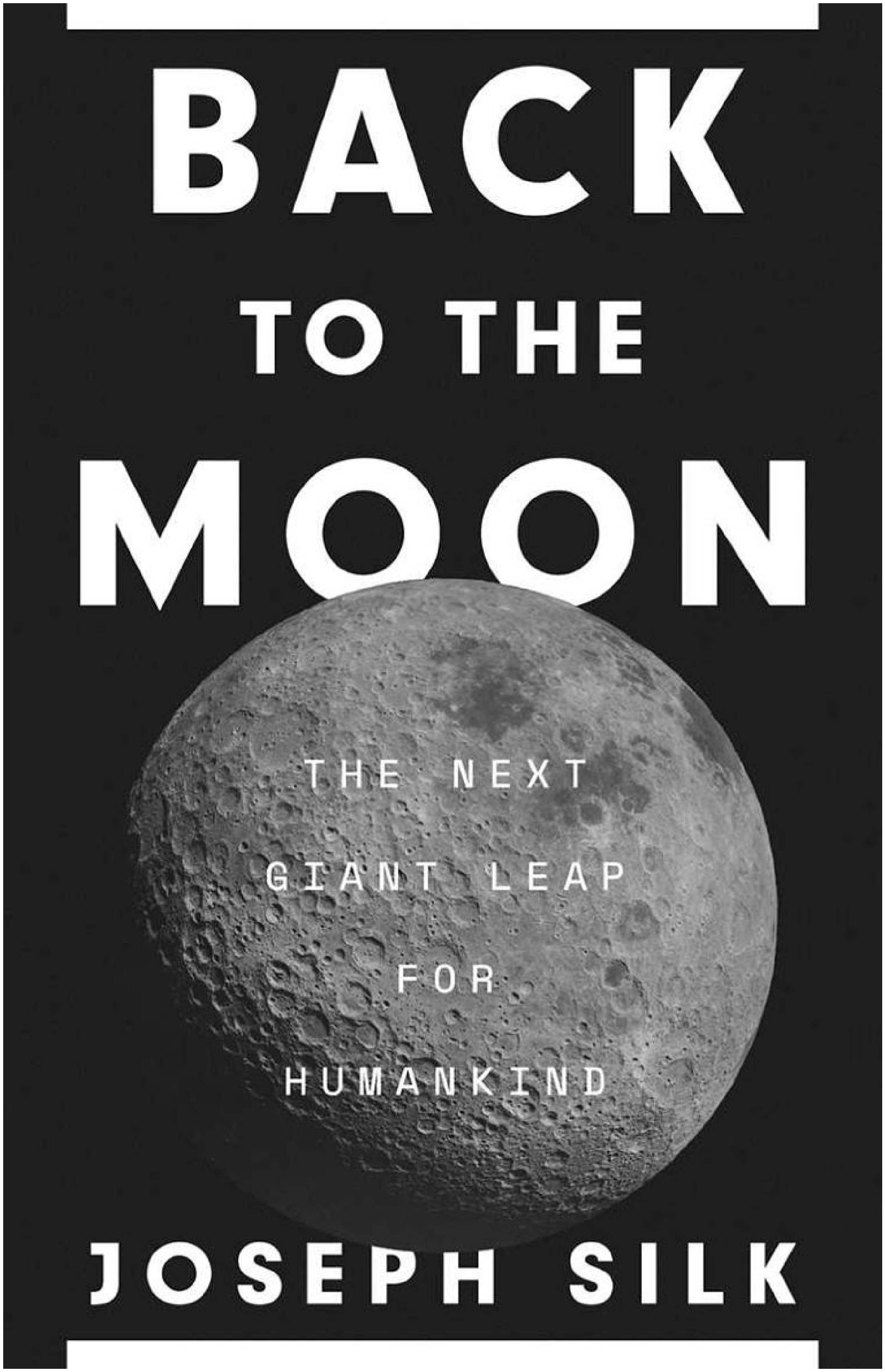
Regulski, curator of a British Museum exhibition. In this companion book, expert contributors vitalize two centuries of Egyptology with erudition and entrancing illustrations.


"Well, Doc, You're In"
FREEMAN DYSON'S
JOURNEY THROUGH
THE UNIVERSE
EDITED BY
DAVID KAISER

“Well, Doc, You’re In”

Ed. David Kaiser *MIT Press* (2022)

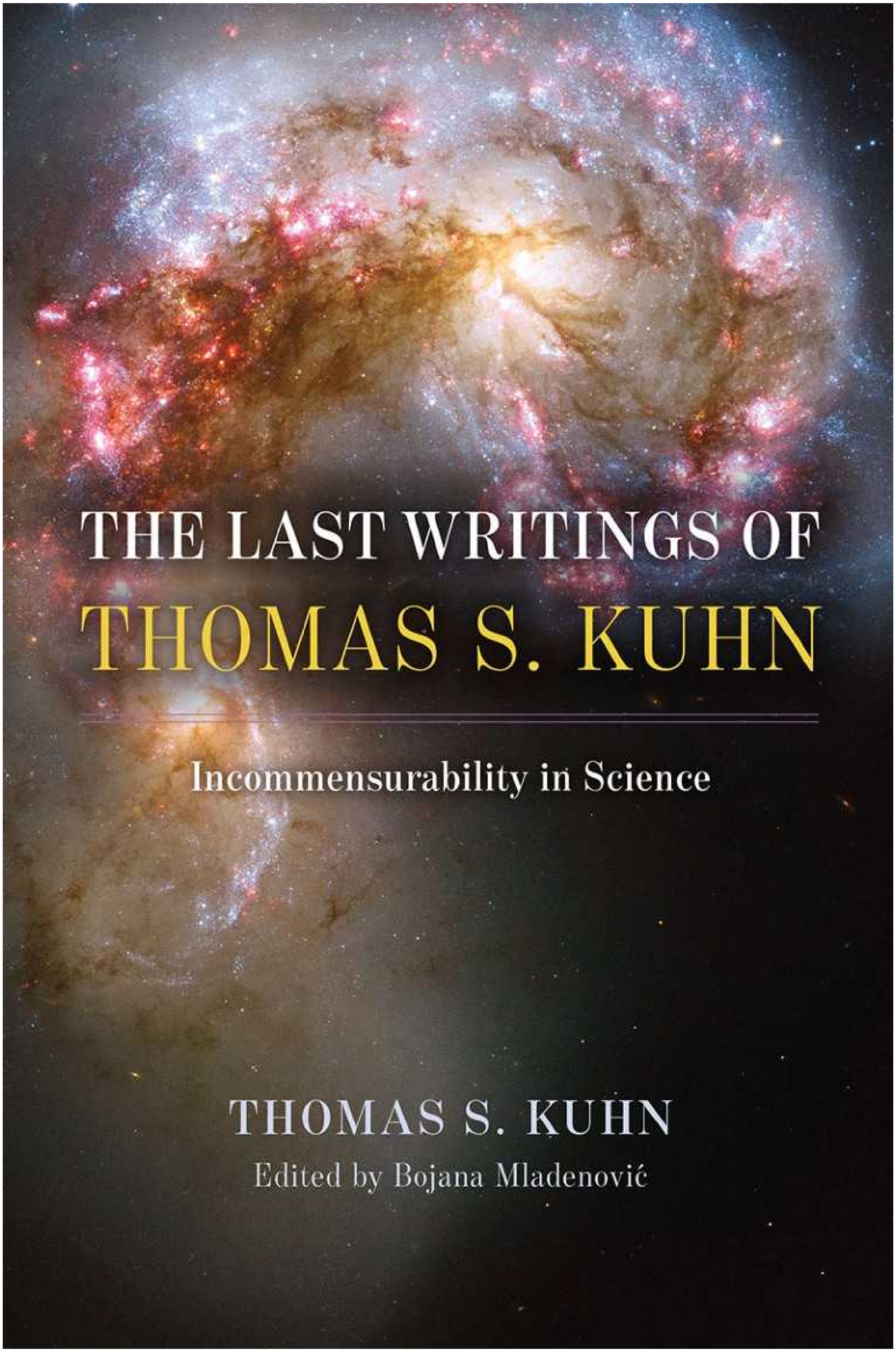
Theoretical physicist Freeman Dyson had a contrarian streak when it came to acceptance into the establishment. After a colleague extolled his work at a 1949 conference, physicist Richard Feynman told Dyson: “Well, Doc, you’re in.” But Dyson proudly never finished his PhD. He was a “determined autodidact” who “developed a lifelong suspicion of organized curricula”, says science historian David Kaiser, editor of this appealing collection of tributes. Hence, perhaps, his early espousal but later denial of the scientific consensus on climate change.


BACK
TO THE
MOON
THE NEXT
GIANT LEAP
FOR
HUMANKIND
JOSEPH SILK

Back to the Moon

Joseph Silk *Princeton Univ. Press* (2022)

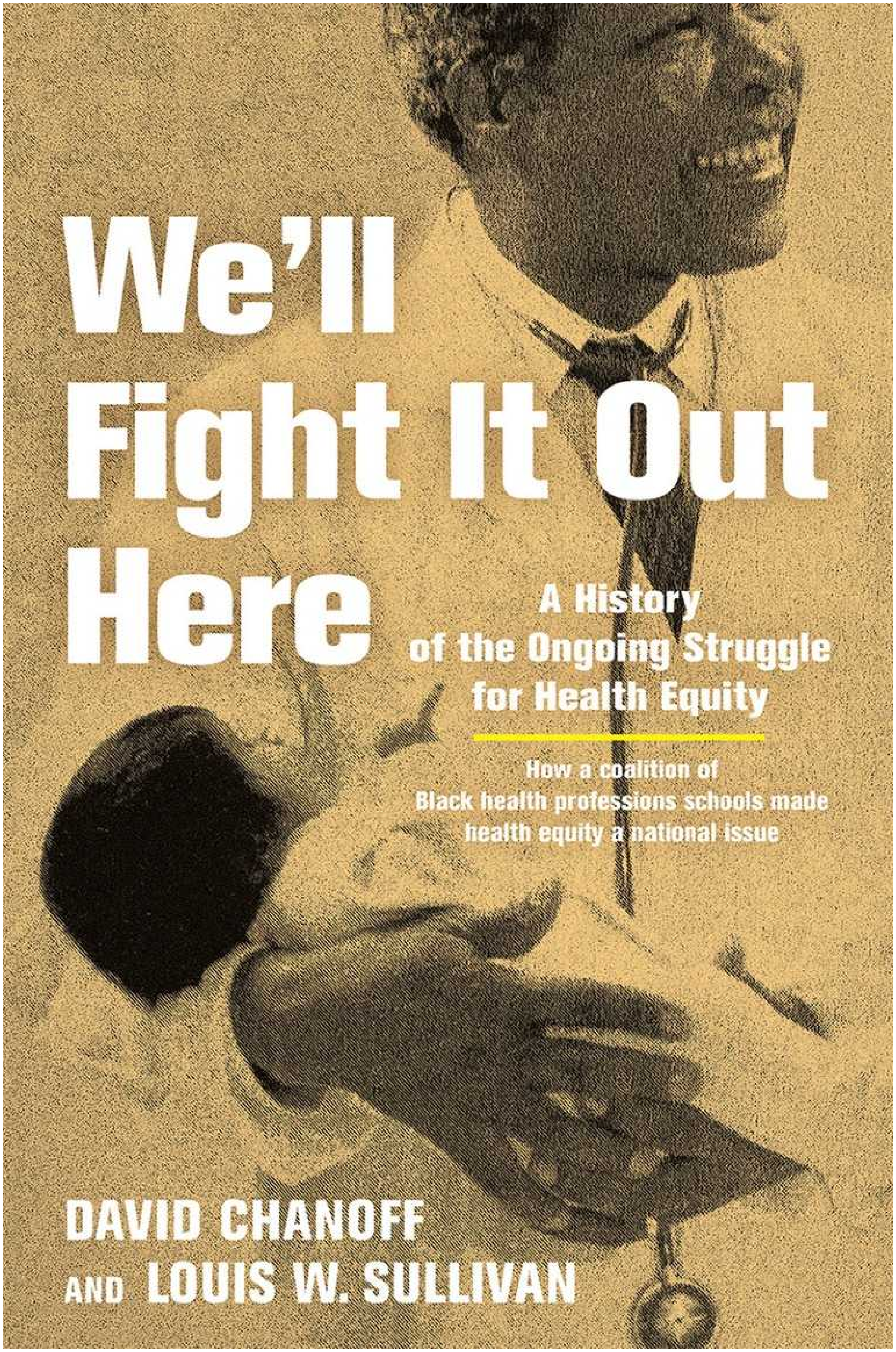
Astrophysicist Joseph Silk won his doctorate in 1968, the year before Apollo 11 landed on the Moon. In this book setting out his vision for the next half-century, he argues with passion for further lunar exploration as soon as possible. He posits that only telescopes on the Moon can realistically probe the origins of the Universe and the possibility of extraterrestrial life. But he recognizes the commercial and environmental risks. International legal treaties, he says, must forbid “free-for-all lunar exploitation reminiscent of the Wild West”.


THE LAST WRITINGS OF
THOMAS S. KUHN
Incommensurability in Science
THOMAS S. KUHN
Edited by Bojana Mladenović

The Last Writings of Thomas S. Kuhn

Ed. Bojana Mladenović *Univ. Chicago Press* (2022)

The Structure of Scientific Revolutions (1962) by philosopher of science Thomas Kuhn is "indispensable reading for every well-educated person", writes philosopher Bojana Mladenović in her introduction to this collection. She presents unpublished drafts of a reworking of *Structure*'s philosophical framework, with the texts of two lectures not previously published in English. Together, these explore whether historians can understand past scientific paradigms, even though these are incommensurable with present science.


We'll
Fight It Out
Here
A History
of the Ongoing Struggle
for Health Equity
How a coalition of
Black health professions schools made
health equity a national issue
DAVID CHANOFF
AND LOUIS W. SULLIVAN

We’ll Fight It Out Here

David Chanoff & Louis W. Sullivan *Johns Hopkins Univ. Press* (2022)

In the first year of the COVID-19 pandemic, the death rate for African Americans was 270% above that for the white population, and the hospitalization rate was 370% higher, note writer David Chanoff and physician Louis Sullivan. Their history of this inequity begins with health during slavery, and focuses on the Association of Minority Health Professions Schools, co-founded by Sullivan in 1976. The group gained national political influence, and Sullivan became head of the US Department of Health and Human Services in 1989.

Nature **614**, 28 (2023)

doi: https://doi.org/10.1038/d41586-022-04569-5

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-022-04569-5

| Section menu | Main menu |

Opinion

- **Alexander Müller (1927–2023)** [24 Jan 2023]
  • Physicist and co-discoverer of high-temperature superconductivity.
- **Could Africa be the future for genomics research?** [31 Jan 2023]
  • Funds for a major genomics programme in Africa will run dry this year. A chance to address global inequity in health-related genomics by building on the success of this initiative must not be missed.

- OBITUARY
- 24 January 2023

Alexander Müller (1927–2023)

Physicist and co-discoverer of high-temperature superconductivity.

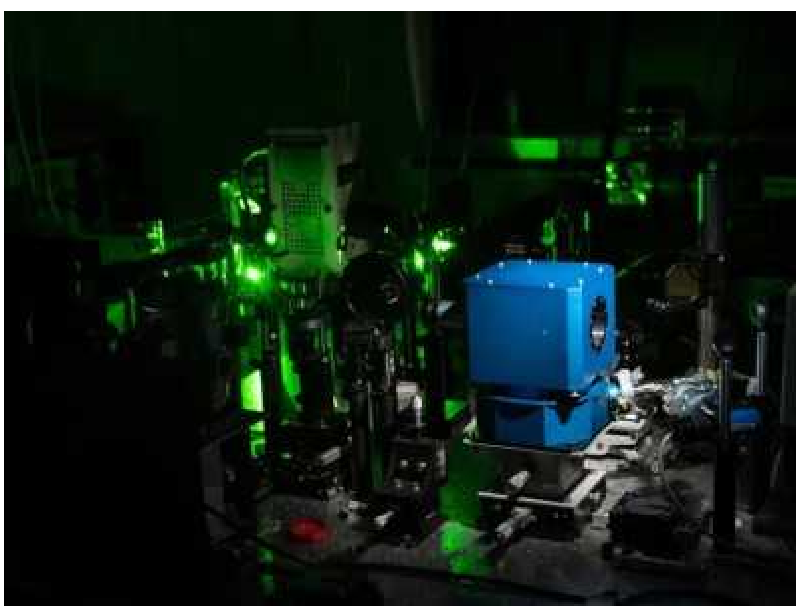
- Joseph D. Martin



Credit: Sueddeutsche Zeitung Photo/Alamy

In 1986, Alexander Müller and his student Georg Bednorz discovered high-temperature (high-T_c) superconductivity. Only a year later, their discovery earned them the 1987 Nobel Prize in Physics. Superconductivity, in which all electrical resistance disappears below a critical temperature, had been known since 1911, but mainly in metals and alloys cooled to within a degree or two of absolute zero. Müller, who has died aged 95, demonstrated the transition to superconductivity at a temperature of 35 K (35 °C above absolute zero) in a class of ceramic known as cuprates, raising the tantalizing possibility of frictionless power distribution, levitating trains and other applications.

The physics world, initially sceptical, was gripped by excitement. At the American Physical Society's March 1987 meeting, later dubbed the 'Woodstock of physics', in reference to the 1969 music festival, more than 2,000 physicists crammed into a hotel ballroom to hear a panel discuss the new superconductors. By this time, others had found related materials with transition temperatures close to 100 K. High-T_c superconductivity has since been used in magnetic resonance imaging systems and in other specialized applications — yet theorists continue to seek a satisfactory explanation for how it works.

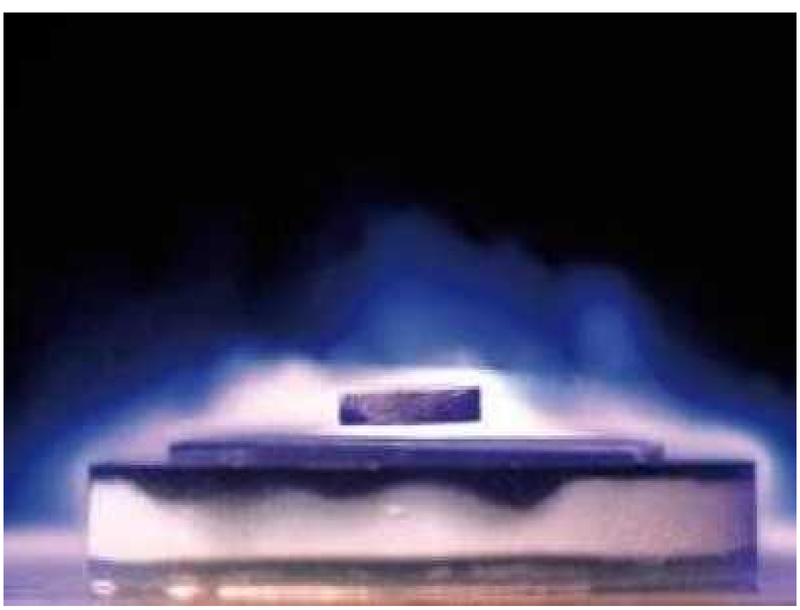


First room-temperature superconductor excites — and baffles — scientists

Born in Basel, Switzerland, in 1927, Müller caught the science bug in childhood, when his mother gave him a vacuum-tube radio kit. At the Swiss Federal Institute of Technology in Zurich, he attended seminars given by the Nobel-prizewinning theoretical physicist Wolfgang Pauli. For his doctoral thesis under Georg Busch, he used electron paramagnetic resonance spectroscopy to study the electrical and magnetic properties of a variety of solids.

In 1958, Müller moved to the Geneva branch of the Battelle Institute, a non-profit industrial research and development organization headquartered in Columbus, Ohio. He studied radiation damage in graphite, in response to the 1957 accident at the Windscale nuclear reactor in the United Kingdom. Partly owing to damage in its graphite moderator, the reactor had overheated and caught fire, releasing a radioactive plume across Europe. Müller's work on neutron-irradiated graphite, published in *Physical Review* in 1961 (K. A. Müller *Phys. Rev.* **123**, 1550–1552; 1961), caught the attention of the director of IBM Research Zurich in Rüschlikon. He lured Müller away from Geneva in 1963 to work on solid-state problems relevant to IBM's investment in digital computing.

Müller spent the rest of his career at IBM Rüschlikon. He established himself as a world-leading expert on perovskites — oxides with a distinctive, symmetrical crystal structure that makes them suitable for a wide range of engineering applications, including microelectronics and solar cells.



High-temperature superconductivity at 25: Still in suspense

Müller took a parallel appointment as a professor at the University of Zurich in 1970, and became director of the physics division at IBM Rüschlikon in 1973. During a two-year secondment to IBM's Thomas J. Watson Research Center in Yorktown Heights, New York, in 1978, he seized on superconductivity as a promising research area — made more tractable by emerging work in critical phenomena, such as phase transitions. After his return to Zurich, he was named an IBM Fellow and given freedom to pursue independent research. He took up high-T_c superconductivity in earnest, and, in 1983, began his collaboration with Bednorz, who shared his fascination with perovskites.

The pair began looking for superconductivity in perovskite-like ceramics, driven by theoretical and experimental hints, but also, as Müller recalled, by an "atavistic type of feeling" of their suitability. In 1986, they observed superconductivity in lanthanum barium copper oxide with a transition temperature of around 35 K, more than 10 K above the previous highest transition temperature.

Cuprates and other high-T_c superconductors are more difficult to work with than are metal alloy low-T_c superconductors, and they have few practical applications as yet. They also do not obey the Bardeen–Cooper–Schrieffer theory, which describes how electrons interact with phonons (packets of vibrational energy) in conventional superconductors. Candidate theories emerged swiftly after 1986, but today high-T_c superconductivity remains perhaps the most notable unsolved problem in condensed-matter theory.



Superconductivity two decades on

Already aged 60 when he received the Nobel prize, Müller continued his work on superconductivity, as an IBM fellow until 1992 and at the University of Zurich until 1994, when he was awarded emeritus status. His research, which he carried on well into his eighties, shed further light on the behaviour of high-T_c superconductors, and placed major constraints on candidate theories of the phenomenon.

Müller's approach to science reflected the expedience of industry, where the distinction between basic and applied research holds little sway. He drew ideas from anywhere he could find them, and quickly abandoned approaches that he found too constraining. He was equally happy to take inspiration from unconventional sources. He recalled being drawn to perovskites because their symmetrical structure reminded him of the mandalas of Eastern religions, to which he attributed a deep, almost spiritual significance.

Alex was an avid sportsman who skied into his eighties and preferred the slopes to the spotlight. Like his nine-year-old self building the radio, he approached physics with the attitude that he had to find out for himself. He flouted conventions that cast doubt on his instincts about superconductivity, declaring, "I just don't talk to the theoreticians. They just held me back."

The fruits of his curiosity continue to define a significant area of condensed-matter physics and its applications.

Nature **614**, 29 (2023)

doi: https://doi.org/10.1038/d41586-023-00206-x

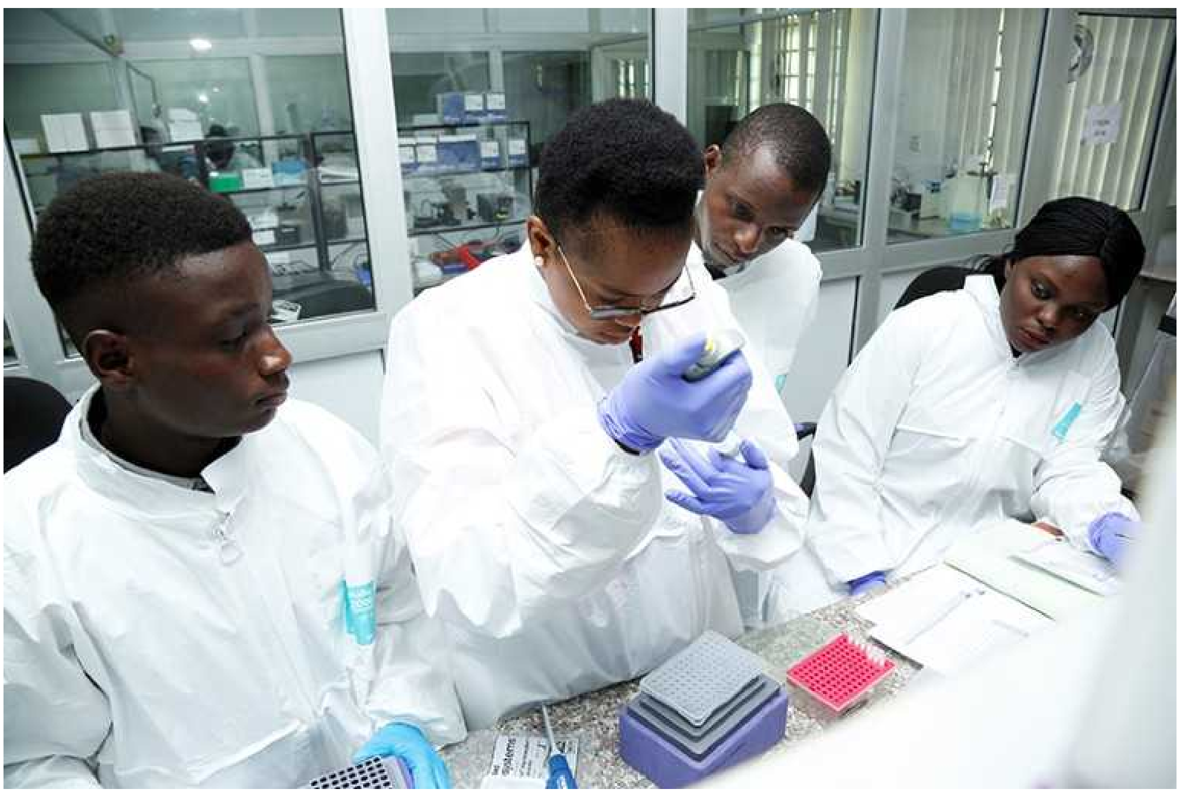
This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00206-x

- COMMENT
- 31 January 2023

Could Africa be the future for genomics research?

Funds for a major genomics programme in Africa will run dry this year. A chance to address global inequity in health-related genomics by building on the success of this initiative must not be missed.

- Zané Lombard &
- Guida Landouré



Scientists at the African Centre of Excellence for Genomics of Infectious Diseases in Ede, Nigeria, prepare samples for DNA sequencing.Credit:

ACEGID

In 2020, an analysis of 426 African genomes, involving researchers from 15 African countries, uncovered 3 million new variants in the human genome[1]. The discovery contributed to the development of a tool that enables researchers to identify genetic associations specifically in African populations — the Infinium H3Africa Consortium genotyping array, produced by the US biotechnology firm Illumina.

Although various enterprises have supported cutting-edge human genomics in Africa, the Human Heredity and Health in Africa (H3Africa) initiative[2], which supported this work, has probably contributed the most in terms of infrastructure and training. The US$176-million programme began in 2010, funded by the US National Institutes of Health (NIH) and the UK biomedical charity Wellcome (in partnership with the African Society of Human Genetics). Projects have ranged from population-based genomic studies of common disorders, such as heart disease, to investigations of infectious diseases, such as COVID-19. Together, some 51 projects, all led by African scientists and involving researchers from more than 30 African countries, have resulted in 50,000 samples being genotyped and nearly 700 papers being published.

Thanks to H3Africa and other genomics initiatives, such as the Nigerian 100K Genome Project[3], African genomics is now poised to improve the health of millions of people worldwide, including those across the continent and the African diaspora. But building on the discoveries made so far — and especially applying findings to the clinic — will require several systemic changes, including a major shift in how genomics research in Africa is funded.

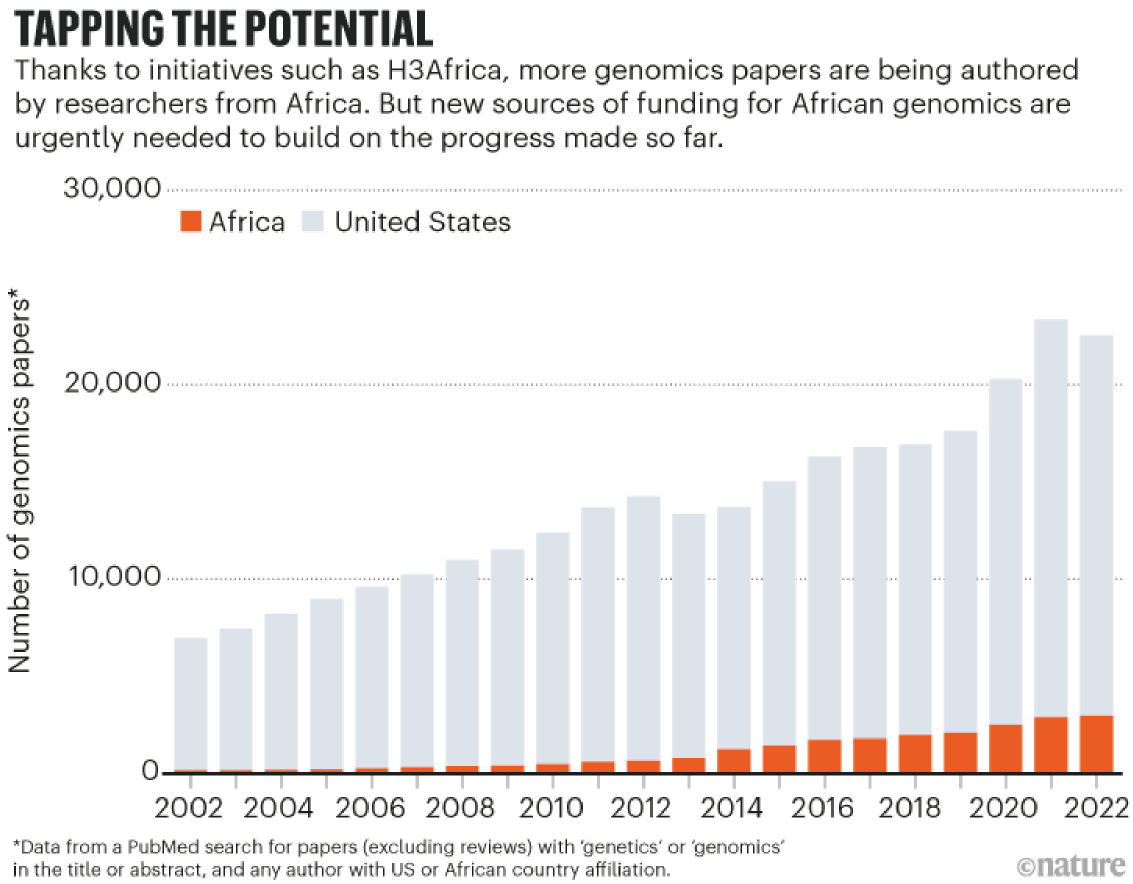
All remaining projects supported by the H3Africa initiative are expected to wrap up this year. (Although funding formally ended in June 2022, some H3Africa grant recipients were able to obtain extensions because of disruption from the COVID-19 pandemic.) Here, we lay out what is needed to ensure that investment in genomics in Africa is not just sustained in a post-H3Africa world but expanded. In our view, Africa could become the birthplace for a new kind of genomics — one that brings better health to all.

What H3Africa has achieved

As of 2021, nearly 86% of participants in genome-wide association studies (GWAS) worldwide were of European descent, even though that group makes up only 16% of the world's population[4]. (Such studies screen the genomes of thousands of people to establish whether a particular genetic variant is associated with a trait of interest.) This bias means that precision-medicine tools, such as polygenic risk scores — which use genetic data to predict a person's risk of developing a certain disease — are much more accurate for people of European ancestry[5].

H3Africa has helped to address this inequity in at least five ways.

First, H3Africa has made it easier for Africans to pursue genomics on the continent (see 'Tapping the potential'). This has been made possible through establishing sequencing facilities, including in South Africa, Gambia, Nigeria, Ethiopia, Uganda and Botswana, and setting up training programmes, for example in bioinformatics[6]. Because African institutions were made the direct recipients of grants in the H3Africa programme, Africans have been able to design projects, training and infrastructure according to the needs of their own countries.



Source: PubMed/Analysis by Z. Lombard

Second, H3Africa has increased the profile of African genomics globally. In part, this is thanks to researchers and organizations funded by H3Africa partnering with international bodies, such as the Heads of International Research Organizations (HIRO — composed of the directors of those organizations, including major funders such as the NIH and the Bill & Melinda Gates Foundation).

Third, H3Africa has enhanced education and awareness about genomics across the continent — principally by requiring that all projects involving participants include some amount of community engagement. For example, Genome Adventures, an initiative of an H3Africa-funded network of organizations called the Collaborative African Genomics Network, is likely to have contributed to the high retention rate of participants in the network's research project. In this case, workshops, comic books and social media

were used to educate community stakeholders and the public about genomics and biomedical research.

Fourth, H3Africa has enabled an unprecedented degree of scientific collaboration between researchers — both across the continent and with those elsewhere. It achieved this partly by requiring that funded African researchers collaborate with other African scientists. During the COVID-19 pandemic, for example, researchers at the African Centre of Excellence for Genomics of Infectious Diseases (ACEGID) in Ede, Nigeria, trained more than 1,300 geneticists, public-health workers and officials from other African countries in diagnostics, next-generation sequencing and bioinformatics. (ACEGID, established in 2014, has received around 30% of its funding from H3Africa.)



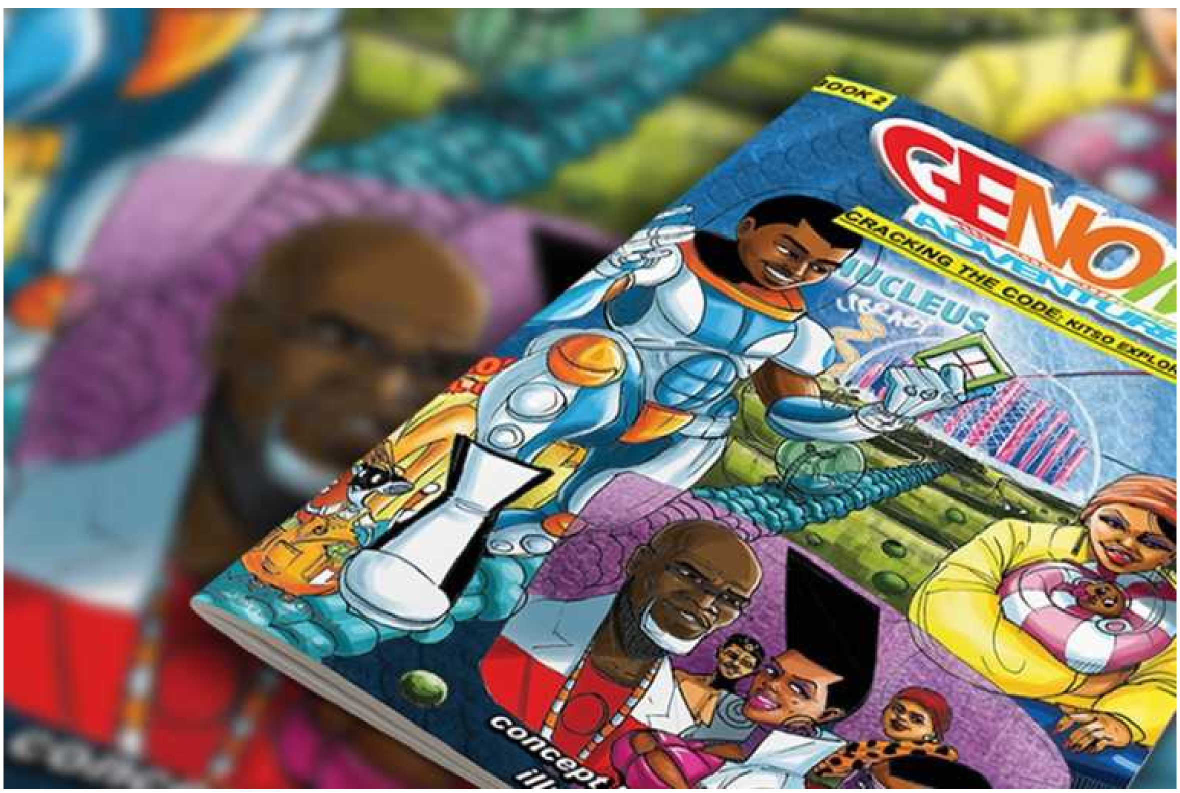
South Africa's San people issue ethics code to scientists

Lastly, by providing ethics guidelines that are tailored to African settings, H3Africa has helped to establish standards and norms around data sharing[7] and the communication of findings to study participants. Although the rules and regulations on this vary greatly across Africa's 54 countries, geneticists across the continent now have a central resource for guidance on informed consent, stigma around genetic diseases, the sharing of benefits from studies and other topics.

Realistic goals

Although most H3Africa projects have not yet translated findings into clinical practice, some have already had real-world impact. For instance, over the past eight years, work at ACEGID has informed public-health responses to Lassa fever, Ebola and COVID-19. Likewise, to improve paediatric care, Solomon Ofori-Acquah, an H3Africa-funded researcher at the University of Pittsburgh in Pennsylvania and at the University of Ghana in Accra, last year established a [genetic screening programme for newborns in Ghana] that was paired with a genetic-counselling training programme.

But will all this progress in African genomics be sustained after H3Africa funding ends? What is needed to ensure that current and future investments translate into improved health for people?



The Collaborative African Genomics Network, part of the H3Africa consortium, produced a set of four Genome Adventures comic books to improve public understanding of genomics in Botswana.Credit: CAfGEN Genome Adventures/Botswana-Baylor

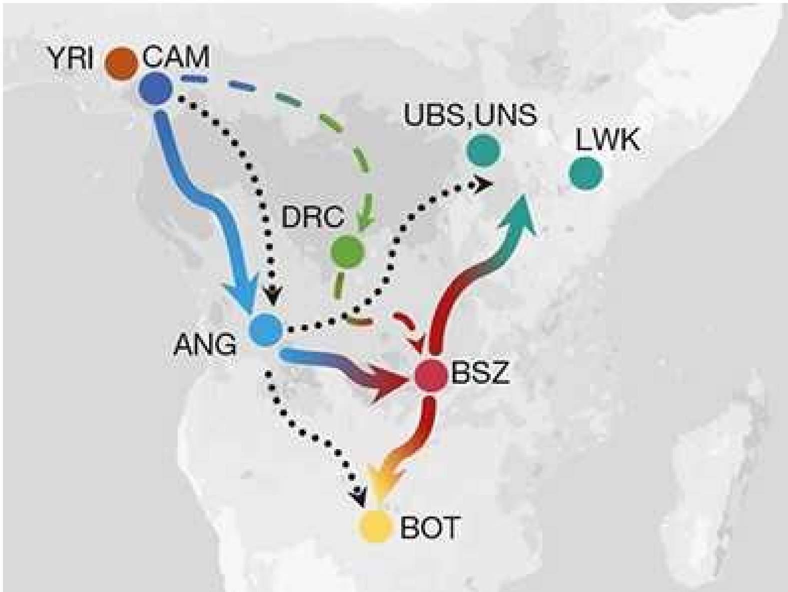
In some areas, it is easy to imagine how real-world genomics applications in Africa could be ramped up. Building a continent-wide early-warning system to detect and track disease outbreaks, for instance, could be achieved largely through scaling up the ACEGID model. An array of such centres across the continent would enable training of African scientists, help to retain African talent in Africa and ensure that standards of practice around African genomic surveillance are developed on the continent.

The Africa Centres for Disease Control and Prevention has already established a surveillance network called the Pathogen Genomics Initiative. This consists of national public health institutes and genomics laboratories, such as ACEGID.

In other areas, the health benefits are farther away. Examples include treatment of common, non-infectious conditions (such as diabetes), largely because of a lack of data, and the treatment of rare diseases, largely because of a lack of genomic medicine services. However, it is becoming clearer what needs to happen next.

Because African genomes have a longer evolutionary history and harbour more variation than the genomes of people with European ancestry, they offer a richer source of variants linked to traits of interest, such as rare developmental disorders. Studies are revealing the incredible potential of African genomes: for developing ways to diagnose diseases across diverse populations; uncovering new therapeutic targets; and identifying genetic markers that indicate how someone might respond to a particular drug.

As an example, throughout the world, a mutation in a gene on chromosome 4 is used to diagnose Huntington's disease. However, another form of the disease exists that is clinically indistinguishable, called Huntington's disease-like 2 (HDL2). This is caused by a mutation in a gene on chromosome 16[8,9]. So far, all cases of HDL2 have been found in people with African ancestry. This suggests that individuals who have Huntington's disease symptoms but do not have the chromosome 4 mutation should be tested for the mutation on chromosome 16.



[Read the paper: High-depth African genomes inform human migration and health]

To deliver the greatest yield per investment when it comes to looking for translatable findings in African genomes, a network of genomic centres needs to be established across Africa. Geneticists operating in such a network should collaborate with those working on multi-year cohort studies to tease apart the effects of genetic and environmental factors, particularly for complex diseases.

One encouraging sign is a funding award to help establish the African Population Cohort Consortium. Last year, the African Population and Health Research Center in Nairobi received funding from Wellcome to develop and co-lead this consortium with the Africa Health Research Institute in Durban, South Africa. It will collate health-surveillance data and biospecimens to provide a resource for large-scale population studies.

Ideally, investment in an African genome initiative at scale would happen alongside the establishment of a pan-African biobank similar to the UK Biobank, which has been periodically collecting health and genetic information from more than 500,000 participants since 2006. H3Africa has already established biorepositories in Uganda, Nigeria and South Africa, which involve the collection of samples linked to the H3Africa genetic data resource. The Genetics of Latin American Diversity (GLAD) Project[10],

which includes genome-wide data on around 54,000 individuals from 39 studies, shows that such a standardized, pan-continental approach is feasible.

Systemic changes needed

In July last year, the World Health Organization (WHO) Science Council — which advises the WHO director-general on high-priority scientific issues — stressed the importance of genomics to future global health in a special report[11]. In our view, heeding the WHO's call and realizing the goals laid out here will require four systemic changes to genomics research in Africa.

Government investment. African governments have been slow to support genomics infrastructure. None of them has delivered on pledges made at the 2007 African Union assembly to spend 1% of gross domestic product on research and development by 2010 (see ref. [12] and go.nature.com/3jt9dss).



Participants in H3Africa's Deciphering Developmental Disorders in Africa research project attend a community Information session in the Democratic Republic of the Congo.Credit: Prince Makay

Convincing political leaders in Africa to invest in genomics has been challenging, in part because it can take decades before pay-offs are realized. The fact that most evidence for the benefits of genomics comes from research on European-ancestry populations in high-income settings also makes it harder for scientists to convince African governments that genomics could help their own people.

Yet as long as researchers rely on grants from outside funders such as the NIH, African governments won't have ownership of projects or be able to set priorities. What's more, sustained government commitment to genomics, which would enable researchers and health-care practitioners to keep working in Africa, is crucial to bringing the benefits of genomic medicine to local populations.

Matched funding schemes, whereby funds are provided by a donor on the condition that the receiver also contributes resources (similar to those implemented by the UK Newton Fund in selected African, Asian and South American countries), could help to shift trends in Africa away from over-reliance on donorship.

Industry buy-in. South Africa houses around one-quarter of all the next-generation sequencing facilities on the continent[13]. For genomic medicine to have an impact on the health of millions, data need to be generated that are of similar quality to those used in the global north, and they must be produced at a similar pace and price.



Two years of COVID-19 in Africa: lessons for the world

More government investment would help with this. Public–private partnerships are also key. Major pharmaceutical companies are increasingly expressing interest in scaling up translational genomics in Africa. At the American Society of Human Genetics annual meeting last October, for example, an emerging consortium of seven firms met with H3Africa investigators to discuss how current fragmented initiatives might be transformed into an integrated public–private partnership.

The responsible engagement of industry — based on principles of African ownership, and the equitable distribution of credit and benefits — could help to resolve crucial needs, such as the lack of maintenance staff for sequencing machines, affordable reagents and reliable supply chains.

Various endeavours have already demonstrated the promise of public–private partnerships. For instance, a genomics centre was established in 2018 in Cape Town, South Africa, thanks to a partnership between the Chinese genomics company BGI Group and the South African Medical Research Council. It was one of the facilities used to detect Omicron SARS-CoV-2 variants in South Africa's wastewater using high-throughput sequencing. Similarly, in 2021, the Illumina and Genetic Alliance, a non-profit organization that advocates research on rare diseases, launched the iHope Genetic Health programme. This $120-million initiative aims to expand

global access to whole-genome sequencing, with more than one-third of the money going to Africa. But such alliances need to be much more widespread.

Genetic services and electronic health records. In the United States, there are 2 medical geneticists and 7 genetic counsellors for every 500,000 people[14]. In South Africa, which offers the most extensive medical genetics services in Africa, both these numbers are less than 0.2 (see go.nature.com/3yjoxpk and ref. [15]).

For people in Africa to benefit from potentially transformative genomic medicine applications, such as emerging gene-based therapies for sickle-cell disease, genetic counselling services are needed in every country. (Even with limited screening, sickle-cell disease is known to affect around 225,000 babies in Africa annually[16]. By our estimate, the disease results in 8–15 deaths per 1,000 children under 5 years old.) Electronic health-record systems are also essential if the use of genomic information in clinical care is to become routine.



A physician in Cameroon discusses prenatal testing with a patient. Such screening for genetic diseases is limited across Africa.Credit: Heiner Heine/imageBROKER/Shutterstock

Equitable science and meaningful engagement of research participants. The legacies of colonialism and scientific racism continue to block opportunities for African scientists to participate equitably in genomics[17]. Research that focuses on the diseases most prevalent in Africa are often led by non-African organizations, for instance. Researchers globally have frequently failed to consider how benefits stemming from their work could be shared equitably, and some genome data have been used unethically[18,19].

Changing this will require a commitment from all stakeholders to promote equitable collaboration with partners in Africa, modelled on the successes of programmes such as H3Africa. A broader array of funders, including major philanthropic organizations, needs to support genomics initiatives on the continent. Priority should be given to proposals for projects that are led by African scientists, that represent African populations and that focus on research questions pertinent to African communities.

To ensure that research is shaped according to the priorities of people living in Africa, engagement with research participants must also be tailored to specific cultures and languages. Some 2,000 languages are spoken in Africa, representing a significant barrier to communication. The Lyfe Languages initiative, a project developed in Western Australia to help overcome a barrier between Aboriginal people and health-care providers, offers one model for how this might be achieved. The project provides Indigenous-language translations of terms often used in clinical-genetics research[20].

The next decade

Some researchers who have been reliant on H3Africa funds are in a strong position to compete for the global pot of money available for genomics research. This includes that offered by the NIH and the US National Science Foundation. Many have been awarded grants to join a new consortium called Data Science for Health Discovery and Innovation in Africa (DS-I Africa). This effort, which is focused on data science and analysis, builds on

H3ABioNet, H3Africa's bioinformatics training programme, and is being funded for up to ten years by the NIH Common Fund Initiative.

Most researchers urgently need African governments and other funders to step up to help them secure a viable future in research on the continent. What's more, government commitment is crucial to securing a sustainable future for personalized medicine and medical genetics services in African countries.

This is an opportune moment for a 'moonshot' enterprise to advance the potential of genomics in Africa. A group of H3Africa investigators and NIH colleagues have come together to propose a programme of Genomics Centres of Excellence. These could absorb multiple genomics projects, scale up training, adopt common standards and, through institutional strengthening, offer greater sustainability than sponsorship of individual projects alone. The plan will be presented later this month at the International Congress of Human Genetics in Cape Town, South Africa.

An incredible amount has been achieved in little more than ten years. But with the right investment from the right stakeholders, Africa could achieve so much more in genomics in the decade ahead — with benefits that reach far beyond the continent.

Nature **614**, 30-33 (2023)

doi: https://doi.org/10.1038/d41586-023-00222-x

A full list of H3Africa Steering Committee members can be found in the Supplementary information.

References

1. Choudhury, A. *et al. Nature* **586**, 741–748 (2020).

2. The H3Africa Consortium *et al. Science* **344**, 1346–1348 (2014).

3. Fatumo, S. *et al. Nature Genet.* **54**, 531–536 (2022).

4. Fatumo, S. *et al. Nature Med.* **28**, 243–250 (2022).

5. Martin, A. R. *et al. Nature Genet.* **51**, 584–591 (2019).

6. Mulder, N. J. *et al. Genome Res* **26**, 271–277 (2016).

7. H3Africa Consortium. *H3Africa Consortium Data Sharing, Access and Release Policy* (H3Africa Consortium, 2020); available at https://go.nature.com/3xrx7ct

8. Krause, A. *et al. Am. J. Med. Genet. B Neuropsychiatr. Genet.* **168**, 573–585 (2015).

9. Margolis, R. L. *et al. Ann. Neurol.* **50**, 373–380 (2001).

10. Borda, V. *et al.* Preprint at bioRxiv https://doi.org/10.1101/2023.01.07.522490 (2023).

11. WHO Science Council. *Accelerating Access to Genomics for Global Health: Promotion, Implementation, Collaboration, and Ethical, Legal, and Social Issues* (World Health Organization, 2022).

12. African Union. *Assembly of the African Union Eighth Ordinary Session: Decisions and Declarations* (African Union, 2007).

13. Inzaule, S. C., Tessema, S. K., Kebede, Y., Ogwell Ouma, A. E. & Nkengasong, J. N. *Lancet Infect. Dis.* **21**, 281–289 (2021).

14. US Government Accountability Office. *Genetic Services: Information on Genetic Counselor and Medical Geneticist Workforces* (US GAO, 2020).

15. Abacan, M. *et al. Eur. J. Hum. Genet.* **27**, 183–197 (2019).

16. Piel, F. B. *et al. Lancet* **381**, 142–151 (2013).

17. Wonkam, A. *et al. Mol. Biol. Cell* **33**, vo2 (2022).

18. Ramsay, M. *Patterns* **3**, 100412 (2022).

19. Schroeder, D., Chatfield, K., Singh, M., Chennells, R. & Herissone-Kelly, P. in *Equitable Research Partnerships: A Global Code of Conduct to Counter Ethics Dumping* 73–87 (Springer, 2019).

20. Köhler, S. *et al. Nucleic Acids Res.* **49**, D1207–D1217 (2021).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00222-x

Work

- **Charting a course to make maths truly universal** [30 Jan 2023]
  • Mathematicians leading decolonization efforts say that building knowledge-sharing partnerships with communities is key.
- **Mapping the Amazon's fish under threat** [30 Jan 2023]
  • Ichthyologist Lucia Rapp Py-Daniel preserves and catalogues the fish of Amazonian waters as their numbers fall because of industrial development.

- CAREER FEATURE
- 30 January 2023

Charting a course to make maths truly universal

Mathematicians leading decolonization efforts say that building knowledge-sharing partnerships with communities is key.

- Rachel Crowell

Find a new job



Illustration by Bex Glendining

Mathematics has the potential to be a great equalizer. Compared with other scientific and technical fields, it requires few expensive physical resources. Sometimes, a whiteboard and a marker are all that's needed.

However, maths is one of the least diverse of the STEM disciplines of science, technology, engineering and mathematics. For instance, the Survey of Earned Doctorates conducted by the US National Science Foundation) showed that, of all 1,915 doctorates awarded in mathematics and statistics in the United States in 2021, none went to people identifying as American Indian or Alaska Native. Just 28 (1.5%) were awarded to Black or African American mathematicians or statisticians, and 33 (1.7%) to researchers who identify as belonging to more than one race.

Maths is built on a modern history of elevating the achievements of one group of people: white men. "Theorems or techniques have names associated to them and most of the time, those names are of nineteenth-century French or German men," such as Georg Cantor, Henri Poincaré and Carl Friedrich Gauss, all of whom were white, says John Parker, head of the mathematical sciences department at Durham University, UK. This means that the accomplishments of people of other genders and races have often been pushed aside, preventing maths from being a level playing field. It has also squelched wider access to rich mathematical ideas developed by people of different backgrounds — such as Chike Obi, James Ezeilo and Adegoke Olubummo, a trio credited by the website Mathematicians of the African Diaspora with having pioneered modern maths research in Nigeria. Another example is Mary Golda Ross, a Cherokee mathematician and engineer who was a founding member of 'Skunk Works', a secretive division of the US aerospace manufacturer Lockheed. There, she developed early designs for space travel and satellites, among other things.



Students and staff at Durham University, UK, discuss how to decolonize the maths curriculum.Credit: Steph Dawson

In 2022, the maths department at Durham launched a decolonization effort that was one of many initiatives aimed at making higher-education maths curricula more inclusive. Mathematicians involved in that and similar initiatives elsewhere explain how they are ushering in a new era of teaching and training for the next generation of mathematicians.

Linking culture to curriculum

Some mathematicians seeking to reform maths curricula taught in North American institutions are doing so by incorporating Indigenous cultural and community concepts. This makes courses relevant to a wider range of students, and boosts their engagement by linking content back to people and cultures.

Edward Doolittle, a mathematician at First Nations University of Canada in Regina, contrasts Indigenous mathematics with the mainstream, global way of teaching maths, in which instructors essentially present the same content regardless of where they're teaching.

Doolittle, who's also a Mohawk person from Six Nations in southern Ontario, says that calculus courses are structured so similarly that he could teach the subject "anywhere the students speak English", and even take over teaching a course midstream.



Decolonizing science toolkit

By contrast, he says that Indigenous mathematics involves getting inside a culture and examining the mathematical thinking in it. He draws a further distinction between Indigenous mathematics and the practice of what he calls "indigenizing mathematics", which, he says, involves searching for cultural examples to use in courses taught in the global version of mathematics.

Indigenizing mathematics tweaks the curriculum when it isn't feasible to fully immerse students in ideas from an Indigenous culture, Doolittle says. "It's very hard, if not impossible, to break out of" the global mathematics system, he notes. By indigenizing mathematics, instructors can stay within the parameters of what they're required to cover while broadening the cultural scope of their curriculum.

Using that approach, "we have respected the knowledge of Indigenous people and are furthering our ties with Indigenous people" while still teaching students core topics, he says. For example, when teaching statistics

courses, Doolittle has discussed a simplified version of the Peach Stone Game, which is based on making wagers and is played in his community. “You can analyse this in terms of a binomial probability distribution,” or the chances of two outcomes over time, he says.

“I would like to encourage many of my colleagues to engage in indigenization efforts, and hopefully to turn up interesting examples from their local area,” Doolittle says.

However, he cautions that “it’s not a matter of opening the book and grabbing an example”. Instead, he says, a thoughtful and carefully planned process “is required to do this authentically and ethically”.

For instance, it’s usually inappropriate “to share knowledge of ceremony, without really strong permissions and ethical approvals in place” from collaborator or elder members of the Indigenous community from which the example is drawn. By telling his students about something similar to the ceremonial Peach Stone Game but not identical, Doolittle avoids sharing information that isn’t meant to be made public.

One guideline to follow is that “if an example is written in a book somewhere, then we can use it”. But, Doolittle says, always double-check with another source that it’s both an accurate and an appropriate example.

Kamuela Yong, an applied mathematician at the University of Hawai‘i–West O‘ahu, stresses the importance of networking, especially if mathematicians don’t belong to the communities from which they hope to draw examples. In 2012, Yong became the first Native Hawaiian to earn a doctorate in applied mathematics, when he obtained his degree from the University of Iowa in Iowa City.



Kamuela Yong teaches maths using traditional Pacific Island navigational techniques.Credit: Kānehūnāmoku Voyaging Academy

In 2019, when Yong began strategizing about indigenizing his pre-calculus course, he wanted to integrate examples from Polynesian ocean-navigation techniques. However, he and his immediate relatives “are not sea people”, he says. “My ancestors voyaged here — someone down the line was on a canoe, but not my parents or my cousins.”

So, engaging with various resources has been a key component of his curriculum overhaul and his accompanying online pre- calculus textbook,

which will eventually be freely available. To begin revamping his curriculum, he identified a few strong examples through online searches.

The Polynesian Voyaging Society, a non-profit research and educational organization based in Honolulu, Hawaii, was a crucial source of information. Part of that organization's mission is to share "the art and science of traditional Polynesian voyaging". Yet as Yong started writing the book, he realized that his own experience didn't give him the knowledge and connections he needed. Without further context, "it didn't feel right" to interweave voyaging examples in his curriculum.

So, Yong engrossed himself in building connections with voyaging communities. He volunteered with various groups and talked to voyaging experts about his project. Their reception of his work was "overwhelmingly supportive", he says.

Once he received that buy-in from people in the Polynesian voyaging community, he felt that he could responsibly include those examples in his courses. Working closely with community members can help mathematicians to avoid crucial mistakes. Yong received constructive feedback that made his examples more realistic. For example, there are certain routes from Fiji to Tonga that aren't travelled, owing to strong ocean currents.



Kamuela Yong teaches students about the Polynesian star compass during an outdoor lecture at the University of Hawai‘i–West O‘ahu.Credit: University of Hawaii – West Oahu

When instructors replace abstract examples that students often find irrelevant with ones “that you can see and experience”, the curriculum entices students to learn, Yong says. Although students might be especially interested in maths examples derived from their own cultural backgrounds, they also benefit from engaging examples that are rooted in other communities.

More students might now be able to reap those benefits, says Yong, because interest in indigenizing university mathematics is growing. In 2020, he co-founded the Indigenous Mathematicians organization. He and the other leaders of the organization plan to pull together related resources on the group’s website. It’s paramount that mathematicians approach indigenizing or decolonizing maths respectfully — but that doesn’t mean it will necessarily happen quickly, Yong notes. “Yes, we want to indigenize it, but we’re going to let it happen a little bit more organically.”

Yong urges other mathematicians to make human connections if they want to introduce maths examples from communities that they don't belong to. Otherwise, "it's appropriation all over again", he notes.

'The next Einstein will be African'

In Africa, too, a unique model of maths higher education is aiming to capitalize on human connections and local cultural context. At the African Institute for Mathematical Sciences (AIMS), which receives public and private funding, decolonizing maths teaching and training means encouraging students from across the continent, and especially women, to learn from leading lecturers, both and African and from elsewhere — and then to put their talents to work for the betterment of African communities.

The institute is a system of five centres of excellence in Cameroon, Ghana, Senegal, South Africa and Rwanda that are designed to deliver the next generation of leading mathematical thinkers on the continent. AIMS's five centres award fully funded master's degrees and doctorates, preparing students for jobs in academia and in industry. AIMS is built around the motto "We believe the next Einstein will be African".



The inaugural cohort of AIMS Ghana's industry-partnership programme, run in cooperation with the European School of Management and Technology in Berlin.Credit: AIMS Ghana

"We always have a Pan-African cohort in each centre," says Sampson Adotey, senior manager of outreach for AIMS in Cape Town, South Africa. Adotey, who is Black, was born and raised in Africa. Each centre hosts, on average, people from 15 African countries, studying and living together in a 24-hour learning environment, he says. "Students from the same country do not stay together — we mix them up." Students are encouraged to leave their countries' geopolitical differences at the door, Adotey says, an approach that fosters unity in the student body.

But it is AIMS's focus on supporting female mathematicians that led Angela Tabiri to describe its programme as "life-transforming." Tabiri, an African mathematician at AIMS in Accra, Ghana, whose research focuses on quantum algebra, first came to AIMS Ghana to pursue her master's degree in

2013. In 2019, she earned her doctorate in mathematics from the University of Glasgow, UK.

Tabiri had always planned to move back to Ghana to be near her family. But once she returned in 2019, her early-career struggles echoed those of many other African mathematicians. Despite her passion for sharing her maths talents at home, and having interviews at universities, the right job didn't come up. "The system is a bit slow here," Tabiri says.

Other African mathematicians who've trained abroad often find that such job-market pressures stop them from returning. Low salaries are one big driver of that talent loss, Adotey notes. African mathematicians often find that job opportunities in their home countries pay a pittance compared with elsewhere.

But Tabiri's AIMS connections created a pathway for her to pursue maths research in Ghana. On one of her return visits, AIMS's president encouraged her to apply for a postdoctoral fellowship, which she held for two years before becoming a research associate.

"It was great for me to have passed through AIMS and be back again to teach," Tabiri says. When students see "a woman of their same colour" who trained at AIMS, earned a doctorate and returned to work there, it's encouraging, she notes. "It makes them believe they can also get a PhD." Out of more than 2,400 AIMS alumni, one-third are women.



Angela Tabiri teaches a free nine-month maths course for secondary-school girls in Ghana.Credit: Angela Tabiri

AIMS also offers a 'mini-MBA' programme and a cooperative master's degree programme. The latter pairs students with industry partners, such as banks and other financial services companies, software companies, development organizations or international research centres, for internships that last from three to six months. These learning experiences prepare them to work in industries such as big data and IT security. And Adotey says that several industry partners have hired alumni as employees.

It was more than chance that enabled Tabiri to return to AIMS Ghana. To create the future Africa that AIMS hopes for — one in which maths expertise is a catalyst for economic prosperity — the organization is using a hybrid approach to recruiting talent.

Faculty members at the centres are hired from African countries, often through partnerships with local universities. AIMS also hosts visiting lecturers from outside Africa who teach courses that range from a few weeks to two months in length. Bringing in outside researchers exposes students to top talent while they continue to expand their roots in Africa's mathematical communities.

Now, Tabiri is cultivating Ghana’s future mathematical talent in another way. She’s the academic manager of AIMS’s Girls in Mathematical Sciences, a free nine-month programme for secondary-school students.

The scheme was founded because too few women were applying to AIMS. It equips girls with free books, a laptop and Internet access. “Some of them haven’t even seen computers,” Tabiri says. Through virtual masterclasses, academic and industry mathematicians discuss “research in a way the students can engage with”.

Each girl is paired with a mentor, and students visit AIMS Ghana during school breaks, with free meals and lodging, to attend mini courses on finance and physics. In the first cohort of students, 35 girls attended from 11 of Ghana’s 16 regions, Tabiri notes. They graduated from the programme in November 2021. “For those who have applied to university, 21 of them are doing STEM courses.”

Tabiri is confident that, in a decade from now, “these girls will be leading in research and innovation in the mathematical sciences”.

Facing misconceptions

The next generation also leads the way in the decolonization efforts at Durham’s maths department — but these haven’t been without bumps in the road. Sometimes, the efforts have led to confusion or political backlash. Just using “that word ‘decolonization’ may mean a lot of opposition that you should be ready for”, says Emmanuel Ogundimu, co-director of Durham’s biostatistics unit, who leads the decolonization initiative. But he thinks that, with patience, this work can create meaningful change.

Durham’s maths decolonization work is part of a wider movement in the United Kingdom to “decolonize the curriculum across all universities and across all subjects”, says Parker, who is white British. He says that departments were asked by the university to think about what they could do.

The Black Lives Matter movement, which accelerated following the murder in 2020 of George Floyd by a police officer in the United States, moved “the whole agenda into the forefront in this system”, says Ogundimu, who hails

from Nigeria and is of Black African descent. Ogundimu and Ric Crossman, a statistician at Durham, selected three undergraduate interns to work on the decolonization project, from a pool of students who applied. Those interns – Harini Pradeep, who is of South Asian descent, and Claire De-Korte and Leah Parry, who are both white – are an integral part of the project.



Decolonizing science toolkit

From the onset, the interns "knew there would be misconceptions" in the outside world about what they're working to accomplish, Ogundimu notes. Sometimes it's even challenging for mathematicians and other researchers to imagine how to decolonize a quantitative subject such as maths, because they're not used to identifying how their curriculum might be affected by colonialist or racist mindsets. Some might initially even view this work as detracting from the rigour of their instruction by focusing on what might seem to be strictly social issues. But mathematicians who have thought deeply about decolonization say that it is essential to creating an academic environment that is welcoming and conducive to the success of maths students and scholars of all backgrounds.

First, the Durham interns asked other UK maths departments about their institutions' decolonization processes. This has revealed that other departments are also struggling to define what the process means for them — is it about rooting racism out of the curriculum, adding more-inclusive

examples, diversifying who is teaching and learning maths, or all of the above?

This year, the students plan to poll other students at Durham about their understanding of decolonization and how it should be accomplished in the maths department, Ogundimu notes. Finally, they are developing a decolonization handbook for faculty members, which will "guide them through the principles of what decolonization is and how that could translate into what we want to achieve overall".

The handbook will focus on three things: "what students are being taught, how it's being taught and who is doing the teaching," Ogundimu says. "What happens in other cultures? How does this impact how mathematics is communicated in general?" he asks. These curriculum changes demonstrate to students that white Westerners aren't the only people doing mathematics.

"These ideas go a long way for getting people engaged," Ogundimu says. "Who says that an example in statistics cannot come from some data set from ancient Egypt?" he asks.

Last year, the department faced scrutiny after some media outlets criticized its decolonization webpage, which was written mainly by Crossman, who is white. One publication even described decolonization attempts as "sinister". The efforts were misrepresented in the media, Ogundimu says.

"Nobody says, 'Middle-aged white men shouldn't be cited.' That is not the thing." Instead, he asks, if these are the only people you have referenced in courses, then what can you do differently to include other people?

The negative coverage shows "just how easy it is to misunderstand this message", Crossman says. And it's "very clear evidence that the message needs to be broadcast more, not less. Because perhaps some people aren't necessarily engaging in good faith with what we're saying. But there's an awful lot of other people who clearly just don't quite understand yet where we're coming from," he notes.

"It's always easy to not do the right thing because you're too busy," Parker says. But he encourages non-minoritized, white mathematicians such as him

to empathize with “people who don’t hear people from their nation or culture mentioned”.

Durham’s senior mathematicians felt that their curriculum-reform process had to be led by the students, because otherwise “we’re in the awful situation of deciding for ourselves what’s best for them”, Crossman says. That, Parker adds, would be at odds with the concept of decolonization, because colonization “was some group of people thinking they knew best for some other group of people”.

Nature **614**, 183-186 (2023)

doi: https://doi.org/10.1038/d41586-023-00223-w

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00223-w

- WHERE I WORK
- 30 January 2023

Mapping the Amazon’s fish under threat

Ichthyologist Lucia Rapp Py-Daniel preserves and catalogues the fish of Amazonian waters as their numbers fall because of industrial development.

- Patricia Maia Noronha



Lucia Rapp Py-Daniel is an ichthyologist at the National Institute for Amazonian Research in Manaus, Brazil. Credit: Mauro Pimentel/AFP via Getty

When I first came to the Amazon from central Brazil in 1978, I was planning to stay just a year, but I was mesmerized by the size of the rainforest's rivers and its biodiversity. I ended up staying longer and earned my master's degree in aquatic biology in 1984 from the National Institute for Amazonian Research (INPA), in Manaus, Brazil. I then went to get my PhD in ecology and evolutionary biology at the University of Arizona in Tucson, and returned to Manaus in 1998 to work as an ichthyologist at INPA.

I was part of the team that started INPA's fish collection in 1978. At the time, most scientific information on Amazonian fish, including specimens, had been collected by researchers and stored at other institutions around the world. Brazilians couldn't easily access any of it. Now, INPA has preserved and catalogued more than 600,000 fish, all of which are available to our graduate students and scientific community.



Women in science

This picture, from last June, was taken at a Manicoré River creek in northwest Brazil during a Greenpeace expedition. I'm holding a bag of small fish, collected using sieves.

Since 2006, the riverside communities on the Manicoré have been advocating for a reserve to protect their land from non-sustainable practices.

They asked Greenpeace to help map the area's biodiversity to bolster their application. Greenpeace in turn invited INPA researchers for its mapping expedition. We spent 20 days collecting and registering the wide range of creatures in the Manicoré's basins.

Besides fires, the Amazon has been hit hard by deforestation and industrial activities. We registered a decline in populations of several fish species after the construction of the hydroelectric complex of Belo Monte — the second-largest in the world — in the Xingu River. These species can thrive only in the oxygenated environment of running rivers and waterfalls, which have been largely destroyed.

Nature **614**, 188 (2023)

doi: https://doi.org/10.1038/d41586-023-00224-9

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00224-9

| Section menu | Main menu |

Research

- **Measurement of the axial vector form factor from antineutrino–proton scattering** [01 Feb 2023]
  • The authors measure the nucleon axial vector form factor, which encodes information on the distribution of the nucleon weak charge, through antineutrino–proton scattering.
- **Field-linked resonances of polar molecules** [01 Feb 2023]
  • A type of universal scattering resonance between ultracold microwave-dressed polar molecules associated with field-linked tetramer bound states in the long-range potential well is observed, providing a general strategy for resonant scattering between ultracold polar molecules.
- **Triassic stem caecilian supports dissorophoid origin of living amphibians** [25 Jan 2023]
  • Analysis of fossils of the oldest known caecilian provide insights into the origin and morphological and functional evolution of caecilians.
- **Spontaneous behaviour is structured by reinforcement without explicit reward** [18 Jan 2023]
  • Photometric recordings and optogenetic manipulation show that dopamine fluctuations in the dorsolateral striatum in mice modulate the use, sequencing and vigour of behavioural modules during spontaneous behaviour.
- **Insulin-regulated serine and lipid metabolism drive peripheral neuropathy** [25 Jan 2023]
  • Serine deficiency can increase small fibre neuropathy in wild-type mice and serine replacement in diabetic mice alleviates diabetic neuropathy, directly linking amino acid metabolism to peripheral nerve disorders.
- **The person-to-person transmission landscape of the gut and oral microbiomes** [18 Jan 2023]
  • Data from more than 9,700 human stool and oral metagenomes has been used to decipher the strain transmission patterns of the human microbiome from mother to infant, within households and within populations.
- **Dendritic cells direct circadian anti-tumour immune responses** [05 Dec 2022]
  • Rhythmic trafficking of dendritic cells to the tumour draining lymph node governs a circadian response of tumour-antigen-specific CD8+ T cells that is dependent on the circadian

expression of the co-stimulatory molecule CD80.

- **[Programming multicellular assembly with synthetic cell adhesion molecules]** [12 Dec 2022]
  • Synthetic cell adhesion molecules yield customized cell–cell interactions with adhesion properties that are similar to native interactions, and offer abilities for cell and tissue engineering and for systematically studying multicellular organization.
- **[Mitochondrial complexome reveals quality-control pathways of protein import]** [25 Jan 2023]
  • An analysis of MitCOM—a comprehensive resource for the identification, organization and interaction of mitochondrial machineries and pathways in yeast—identifies a constitutive pathway for the removal of preproteins.
- **[Visualization of translation and protein biogenesis at the ER membrane]** [25 Jan 2023]
  • Structural studies of the ribosome-associated endoplasmic reticulum translocon complex based on cryo-electron tomography and molecular modelling reveal multiple intermediate states and interactions between the components of the complex and its cofactors.
- **[Principles of mitoribosomal small subunit assembly in eukaryotes]** [08 Dec 2022]
  • The structures of both human and yeast mitochondrial ribosomal small subunits undergoing assembly are uncovered.

- Article
- Open Access
- Published: 01 February 2023

Measurement of the axial vector form factor from antineutrino–proton scattering

- T. Cai ,
- M. L. Moore,
- A. Olivier,
- S. Akhter,
- Z. Ahmad Dar,
- V. Ansari,
- M. V. Ascencio,
- A. Bashyal,
- A. Bercellie,
- M. Betancourt,
- A. Bodek,
- J. L. Bonilla,
- A. Bravar,
- H. Budd,
- G. Caceres,
- M. F. Carneiro,
- G. A. Díaz,
- H. da Motta,
- J. Felix,
- L. Fields,
- A. Filkins,
- R. Fine,
- A. M. Gago,
- H. Gallagher,

- S. M. Gilligan,
- R. Gran,
- E. Granados,
- D. A. Harris,
- S. Henry,
- D. Jena,
- S. Jena,
- J. Kleykamp,
- A. Klustová,
- M. Kordosky,
- D. Last,
- T. Le,
- A. Lozano,
- X.-G. Lu,
- E. Maher,
- S. Manly,
- W. A. Mann,
- C. Mauger,
- K. S. McFarland ,
- B. Messerly,
- J. Miller,
- O. Moreno,
- J. G. Morfín,
- D. Naples,
- J. K. Nelson,
- C. Nguyen,
- V. Paolone,
- G. N. Perdue,
- K.-J. Plows,
- M. A. Ramírez,
- R. D. Ransome,
- H. Ray,
- D. Ruterbories,
- H. Schellman,
- C. J. Solano Salinas,
- H. Su,
- M. Sultana,

- V. S. Syrotenko,
- E. Valencia,
- N. H. Vaughan,
- A. V. Waldron,
- M. O. Wascko,
- C. Wret,
- B. Yaeggy &
- …
- L. Zazueta

Show authors

Nature **volume 614**, pages 48–53 (2023)

- 2429 Accesses
- 116 Altmetric
- Metrics details

Abstract

Scattering of high energy particles from nucleons probes their structure, as was done in the experiments that established the non-zero size of the proton using electron beams[1]. The use of charged leptons as scattering probes enables measuring the distribution of electric charges, which is encoded in the vector form factors of the nucleon[2]. Scattering weakly interacting neutrinos gives the opportunity to measure both vector and axial vector form factors of the nucleon, providing an additional, complementary probe of their structure. The nucleon transition axial form factor, F_A, can be measured from neutrino scattering from free nucleons, $\nu_{\mu}n \to \mu^{-}p$ and \({\bar{\nu }}_{\mu }p\to {\mu }^{+}n\), as a function of the negative four-momentum transfer squared (Q^2). Up to now, $F_A(Q^2)$ has been extracted from the bound nucleons in neutrino–deuterium scattering[3,4,5,6,7,8,9], which requires uncertain nuclear corrections[10]. Here we report the first high-statistics measurement, to our knowledge, of the \({\bar{\nu }}_{\mu }\,p\to {\mu

}^{+}n\) cross-section from the hydrogen atom, using the plastic scintillator target of the MINERvA[11] experiment, extracting F_A from free proton targets and measuring the nucleon axial charge radius, r_A, to be 0.73 ± 0.17 fm. The antineutrino–hydrogen scattering presented here can access the axial form factor without the need for nuclear theory corrections, and enables direct comparisons with the increasingly precise lattice quantum chromodynamics computations[12,13,14,15]. Finally, the tools developed for this analysis and the result presented are substantial advancements in our capabilities to understand the nucleon structure in the weak sector, and also help the current and future neutrino oscillation experiments[16,17,18,19,20] to better constrain neutrino interaction models.

Main

Form factors measured in scattering processes describe the structure of composite objects. They have been thought to be the Fourier transform of charge distributions in the non-relativistic limit of low negative four-momentum transfer squared, Q^2. Although this interpretation is not strictly true[21], the slopes of the form factors at $Q^2 = 0$ provide a measure of the mean-squared radius $\left\langle {r}_{{\rm{A}}}^{2}\right\rangle$ for the particle in the charge species described. Nucleon electric (${G}_{{\rm{E}}}^{{\rm{N}}}$) and magnetic (${G}_{{\rm{M}}}^{{\rm{N}}}$) form factors are precisely measured in electron–nucleon elastic scattering experiments, enabling the radius of the nucleon[22,23] to be inferred. Neutrino scattering measurements yield the analogous axial vector form factor, F_A, which characterizes the weak charge distribution. F_A is also a key input to neutrino oscillation experiments to precisely measure the neutrino oscillation parameters, including CP violation, and to establish mass hierarchy.

Previous measurements of F_A in neutrino scattering were performed by measuring $d\sigma/dQ^2$ in the reaction $\nu_\mu D \rightarrow \mu^- pp$ in deuterium bubble chambers[3,4,5,6,7,8,9]. Even in deuterium nuclei, theoretical assumptions[10] about the Fermi motion of the bound nucleons, the application of the Pauli

exclusion principle in the proton-proton (*pp*) final state and the nuclear wave function are required to extract F_A from these measurements[24].

Previous extractions assumed F_A to follow the dipole form factor, although more flexible models for the form factors are also available[24,25,26]. Although there are many efforts[12,13,15] to calculate F_A for $Q^2 > 0$ from lattice quantum chromodynamics (QCD) with increasing precision, calculations in the Q^2 region above 1 GeV^2 remain imprecise. Pion electroproduction[27] and muon capture measurements on hydrogen[28] may also be interpreted as constraints on F_A, but again with theoretical uncertainties.

In the charged-current elastic (CCE) reaction on the free nucleon $\bar{\nu}_{\mu}{\rm{H}}\to {\mu}^{+}n$, a muon antineutrino elastically scatters off the free proton from the hydrogen atom, turning the neutrino into the more massive positively charged muon μ^+ and the proton into a neutron (see section 'Terminology of the signal process'). This reaction is free from the nuclear theory corrections described above in scattering from deuterium (D) and provides a direct measurement of F_A. It is also a two-body reaction with a nucleon at rest; therefore, the neutrino direction and the final-state μ^+ momentum fully specify the interacting system. The antineutrino energy ($E_{\bar{\nu}}$) and Q^2 are reconstructed under the CCE hypothesis ($Q_{\rm{QE}}^{2}$) of equation (1) and use only the muon energy (E_μ), momentum (p_μ) and angle (θ_μ) of the muon with the neutrino beam. The free proton in the initial state, and the neutron and muon in the final states have masses denoted as M_p, M_n and m_μ, respectively.

$$\begin{array}{l}\hspace{4.99878pt}{E}_{\bar{\nu }}=\frac{{M}_{{\rm{n}}}^{2}-{M}_{{\rm{p}}}^{2}-{m}_{\mu }^{2}+2{M}_{{\rm{p}}}{E}_{\mu }}{2({M}_{{\rm{p}}}-{E}_{\mu }+{p}_{\mu }\cos {\theta }_{\mu })},\\ {Q}_{{\rm{QE}}}^{2}=2{E}_{\bar{\nu }}({E}_{\mu }-{p}_{\mu }\cos {\theta }_{\mu })-{m}_{\mu }^{2},\end{array}$$
(1)

The neutron momentum in components parallel and transverse to the neutrino direction are

$${p}_{{\rm{n}}}^{\parallel }=\frac{{M}_{{\rm{n}}}^{2}+{{p}_{\mu }^{\perp }}^{2}-{({p}_{\mu }^{\parallel }-{E}_{\mu }+{M}_{{\rm{p}}})}^{2}}{2{({p}_{\mu }^{\parallel }-{E}_{\mu }+{M}_{{\rm{p}}})}^{2}},$$
(2)
$${{\bf{p}}}_{{\rm{n}}}^{\perp }=-{{\bf{p}}}_{\mu }^{\perp }.$$
(3)

The neutron in non-CCE background reactions will deviate from the predicted neutron directions due to nuclear effects, different initial state assumptions and final-state mass, making available physics-driven selections to reduce the backgrounds. No statistically significant measurement of the process has been performed so far; the only measurement previous to this work recorded 13 events in a hydrogen bubble chamber [29].

The free nucleon cross-section is given by

$$\frac{{\rm{d}}\sigma }{{\rm{d}}{Q}^{2}}\left(\begin{array}{c}\nu n\to {l}^{-}p\\ \bar{\nu }p\to {l}^{+}n\end{array}\right)=\frac{{M}^{2}{G}_{{\rm{F}}}^{2}{\cos }^{2}{\theta }_{c}}{8{\rm{\pi }}{E}_{\nu }^{2}}\left[A({Q}^{2})\mp B({Q}^{2})\frac{(s-u)}{{M}^{2}}+C({Q}^{2})\frac{{(s-u)}^{2}}{{M}^{4}}\right],$$
(4)
$$\begin{array}{ccc}A({Q}^{2}) & = & \frac{{m}^{2}+{Q}^{2}}{4{M}^{2}}[\left(4+\frac{{Q}^{2}}{{M}^{2}}\right)|{F}_{{\rm{A}}}{|}^{2}-\left(4-\frac{{Q}^{2}}{{M}^{2}}\right)|{F}_{{\rm{V}}}^{1}{|}^{2}\\ & + & \frac{{Q}^{2}}{{M}^{2}}\left(1-\frac{{Q}^{2}}{4{M}^{2}}\right)|\xi {F}_{{\rm{V}}}^{2}{|}^{2}+\frac{4{Q}^{2}}{{M}^{2}}{F}_{{\rm{V}}}^{1}\xi {F}_{{\rm{V}}}^{2}+{\mathcal{O}}\,\left(\frac{{m}^{2}}{{M}^{2}}\right)],\\ B({Q}^{2}) & = & \frac{{Q}^{2}}{{M}^{2}}{F}_{{\rm{A}}}({F}_{{\rm{V}}}^{1}+\xi {F}_{{\rm{V}}}^{2}),\\ C({Q}^{2}) & = & \frac{1}{4}\left(|{F}_{{\rm{A}}}{|}^{2}+|{F}_{{\rm{V}}}^{1}{|}^{2}+\frac{{Q}^{2}}{4{M}^{2}}|\xi {F}_{{\rm{V}}}^{2}{|}^{2}\right)\end{array}$$
(5)

where $M=(M_{\rm n}+M_{\rm p})/2$ is the average nucleon mass, $G_{\rm F}$ is the Fermi coupling constant, $\theta_{\rm c}$ is the Cabibbo angle, E_ν is the neutrino energy, $(s-u)$

$= 4ME_\nu - m^2 - Q^2$ and m is the charged lepton mass[30]. The parameters A, B and C are functions of F_A, the two vector form factors ${F}_{{\rm{V}}}^{1}$ and $\xi {F}_{{\rm{V}}}^{2}$ derived from the proton and neutron electric and magnetic form factors, and a pseudoscalar form factor F_P whose effect is suppressed by a factor of m^2/M^2 in A. The form factors are real assuming T invariance[30] and charge symmetry. The vector form factors have been precisely parameterized from results of electron scattering experiments[25,26,31,32]. The pseudoscalar form factor is predicted from the axial form factor[33]. The cross-section reported in this paper is a convolution between the free nucleon cross-section and the wide-band neutrino flux[34], with restrictions on the muon kinematics due to the detector geometry. F_A can be derived from the restricted and flux-convolved cross-section given the other form factors.

The nucleon axial radius may be found from the slope of the small Q^2 expansion of F_A at $Q^2 = 0$:

$$\begin{array}{l}{F}_{{\rm{A}}}({Q}^{2})={F}_{{\rm{A}}}(0)\left(1-\frac{\left\langle {r}_{{\rm{A}}}^{2}\right\rangle }{3!}{Q}^{2}+\frac{\left\langle {r}_{{\rm{A}}}^{4}\right\rangle }{5!}{Q}^{4}+\ldots \right),\\ {\left.\frac{1}{{F}_{{\rm{A}}}(0)}\frac{{\rm{d}}{F}_{{\rm{A}}}}{{\rm{d}}{Q}^{2}}\right|}_{{Q}^{2}=0}=-\frac{1}{6}\left\langle {{r}_{{\rm{A}}}}^{2}\right\rangle \end{array} \quad (6)$$

where $F_A(0) = -1.2723 \pm 0.0023$ is derived from neutron decay measurements[35]. We define ${r}_{{\rm{A}}}\equiv \sqrt{\left\langle {{r}_{{\rm{A}}}}^{2}\right\rangle }$.

Experiment description

The CCE process on hydrogen was measured with the MINERvA[11] detector in a ${\bar{\nu }}_{\mu }$ beam produced at the NuMI neutrino beamline[34] at Fermilab with an average energy of 5.4 GeV. Antineutrino

interactions are selected by requiring a μ^+ and detected neutron signatures in the MINERvA detector.

MINERvA is a segmented scintillator detector with hexagonal planes constructed from strips of triangular cross-section assembled into planes perpendicular to the $\bar{\nu}_{\mu}$ beam, and is described in more detail in the Methods. This analysis reconstructs neutrino interactions in the active tracker region of the detector, which consists of the scintillator. The scintillator strips point in either the vertical (X) or one of the ±60° (U,V) directions. This region is fully active, consisting of 128 tracker planes stacked in alternating patterns of $XUXV$. The alternating orientation enables extraction of a three-dimensional position from the strips when charged particles traverse two or more consecutive planes. Muons produced from charged-current neutrino interactions in the MINERvA detector may exit from the rear and enter the MINOS near detector (ND)[36], which is located immediately downstream of the MINERvA detector. The MINOS ND is a fully magnetized scintillator and steel detector that determines the muon's charge and momentum by measuring its curvature and range.

Only muons in the energy range $1.5\ \text{GeV} < E_\mu < 20\ \text{GeV}$ with an opening angle $\theta_\mu < 20°$ with respect to the neutrino direction are selected because they can be efficiently measured by the MINOS ND. The vertex is defined to be the beginning of the muon track. Energy deposits from other charged particles, such as protons and $\pi^\pm$, can be reconstructed into tracks if they span at least four planes. Photon pairs from π^0s can be reconstructed from their electromagnetic showers.

Although neutrons are not directly observable from ionization as charged particles, they produce secondary particles with observable energy deposits when they elastically, quasi-elastically or inelastically scatter in the detector. The dominant interactions produce low-energy protons, which can be observed[37]. Neutrons also scatter undetectably, for example by inelastically knocking out neutrons from carbon nuclei or elastically scattering from carbon, in ways that change the neutron direction and energy. Monte Carlo simulation studies of single-neutron transport in the MINERvA detector show that the angle between the reconstructed and true neutron directions

follows the sum of two exponential distributions, with 68% of the candidates within 12°.

Cross-section extraction

The CCE cross-section is measured in bins of Q^2. Control samples, events with neutrons pointing away from the predicted direction, provide data-driven constraints on the background models as a function of Q^2. The reconstructed neutron directions from the signal events centre around the predicted direction from the μ^+ reconstruction with deviations due to the angular resolution. Additional nuclear effects alter the initial neutron directions when the neutron is produced by the charged-current quasi-elastic (CCQE) neutrino interaction on a bound proton in the carbon nucleus. Fermi motion in the carbon nucleus imparts each bound nucleon with a random initial momentum resulting in the neutron direction further deviating from the two-body calculation. Although the CCQE cross-section is a function of F_A as well, measurements of electromagnetic form factors even in nuclei as light as ^{4}He have shown that nuclear effects obscure the relationship and make extractions of F_A from measurements CCQE on ^{12}C (refs. [38,39,40]) susceptible to uncertain nuclear physics. Additionally, multinucleon knockout, such as two-particle-two-hole (2p2h) reactions, and secondary interactions of outgoing neutrons before exiting the nucleus are also present in carbon. The latter phenomena, collectively termed the final-state interactions (FSI), can change the direction and energy of the neutron, and can produce additional final-state particles, including pions produced through the excitation and subsequent decay of nucleons in the nucleus.

On the basis of the detected final-state particles, events are divided into those with only nucleons in the final state (QELike) and those with mesons present (non-QELike). Although the CCE signal is an exclusive subset of QELike, the carbon CCQE, 2p2h and resonant pion production events may experience FSI and land in either category. Both the signal and the background processes are simulated using a realistic Monte Carlo simulation of the detector based on the GEANT4 (ref. [41]) simulation toolkit. The input model for neutrino-nucleus interactions is based on GENIE[42] with theory and data-driven modifications[43,44,45,46,47,48]. Finally, the Fermi gas initial state model used by GENIE has been reweighted into a spectral function (SF)[49,50] for a more realistic description of nucleus initial states[51].

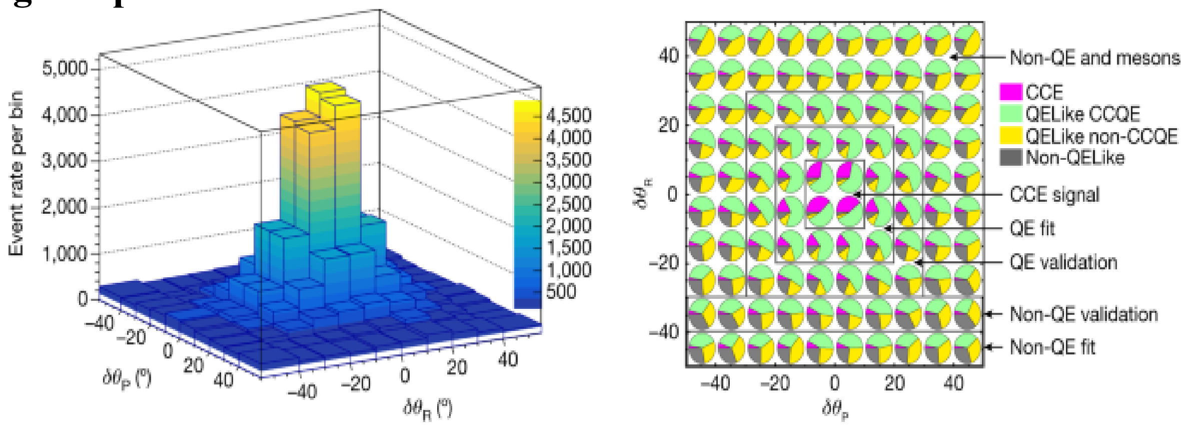
Samples with predominantly QELike(non-QELike) events are selected by requiring ≤1(>1) energetic shower(s) in the detector. Each sample is subdivided according to the neutron candidates’ opening angles to the predicted neutron direction on hydrogen to constrain the QELike and non-QELike backgrounds due to carbon. An alternative coordinate system can be defined as follows:

$$\begin{array}{l}\,{\hat{{\bf{z}}}}^{{\prime} }=\hat{{\bf{t}}},\,{\hat{{\bf{x}}}}^{{\prime} }={\hat{{\bf{p}}}}_{{\boldsymbol{\nu }}}\times {\hat{{\bf{p}}}}_{{\boldsymbol{\mu }}},\,{\hat{{\bf{y}}}}^{{\prime} }={\hat{{\bf{z}}}}^{{\prime} }\times {\hat{{\bf{x}}}}^{{\prime} },\\ \delta {\theta }_{{\rm{P}}}=\arctan ((\hat{{\bf{n}}}\cdot {\hat{{\bf{x}}}}^{{\prime} })/(\hat{{\bf{n}}}\cdot {\hat{{\bf{z}}}}^{{\prime} })),\\ \delta {\theta }_{{\rm{R}}}=\arctan ((\hat{{\bf{n}}}\cdot {\hat{{\bf{y}}}}^{{\prime} })/(\hat{{\bf{n}}}\cdot {\hat{{\bf{z}}}}^{{\prime} })),\end{array}$$
(7)

where the predicted neutron direction $\hat{t}$, together with the neutrino direction ${\hat{p}}_{\nu }$ and measured muon direction ${\hat{p}}_{\mu }$, set up an alternative coordinate system $({\hat{x}}^{{\prime} },{\hat{y}}^{{\prime} },{\hat{z}}^{{\prime} })$. The angular variables $\delta\theta_R$ and $\delta\theta_P$ describe the projected angles between the measured neutron candidate direction $\hat{n}$ and $\hat{t}$ in the neutrino-muon reaction plane and the orthogonal plane intersecting ${\hat{z}}^{{\prime} }$.

Figure 1 shows the relative contributions of the signal CCE and various backgrounds in each $\delta\theta_P$–$\delta\theta_R$ bin in the QELike sample. Although most neutrons from the CCE interaction are reconstructed in the signal region, some undergo rescattering before they are detected, and are measured in non-signal regions. CCQE interaction is the dominant background in the signal region and the regions immediately surrounding it. Fermi momentum and FSI give CCQE events a broader distribution across $\delta\theta_R$ and $\delta\theta_P$. Other interactions have a much broader distribution and are the main background at large $\delta\theta_P$ and $\delta\theta_R$, but still have a small presence in the signal region.

Fig. 1: Data rate and the predicted fractional interaction types in the angular plane.

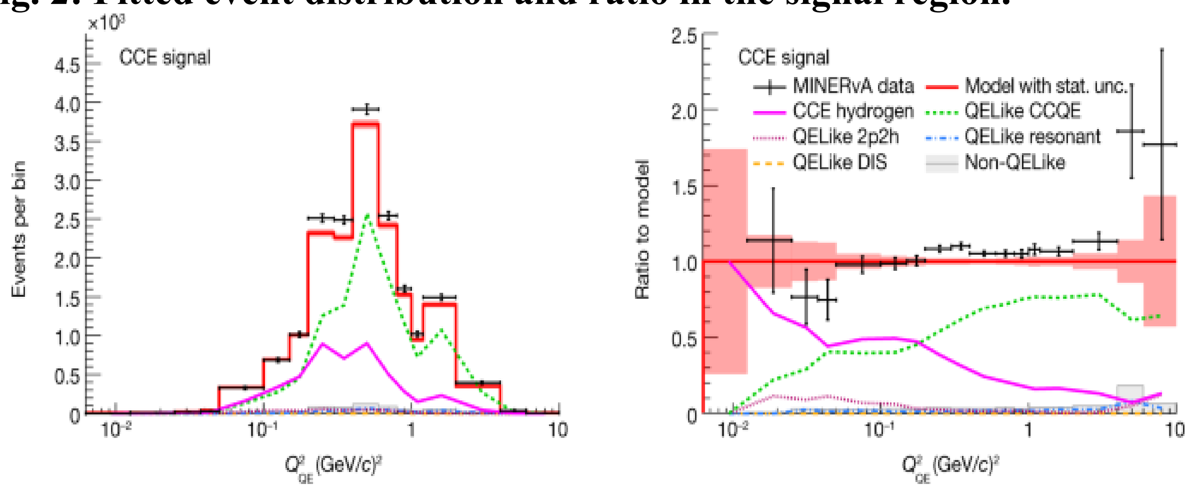


Left, Data event rate in $\delta\theta_P$–$\delta\theta_R$ plane. The visible bins in positive $\delta\theta_P$ are representative of the event rate as the event rate is nearly symmetrical in $\delta\theta_P$. Right, The projected Monte Carlo event fraction in different angular regions. CCE hydrogen (pink), QELike CCQE carbon (green), QELike other (2p2h, resonant, yellow and the non-QELike background (grey) all appear in the event selection. The CCE signal region is between $-10° < \delta\theta_R/\delta\theta_P < 10°$. Control regions are defined to measure events in regions where QE contributions, non-QE contributions and a mixture of non-QE and meson events are dominant. Two validation regions for the QE and non-QE contributions are defined to assess the background estimations. The background levels shown are constrained by the fit to the data as described in the text.

To isolate CCE interactions, the background events are subtracted on the basis of a Monte Carlo simulation. The angular distributions are predicted by the Monte Carlo, but the data determine the Q^2-dependent normalization. The normalization is determined by fitting the 'QE' region, which is dominated by CCQE, together with the 'non-QE' and the 'non-QE and mesons' regions, which have a large fraction of non-CCQE interactions. The normalization as a function of Q^2 of each interaction channel is adjusted to provide the best fit for the three regions. The normalizations are validated by comparing to event rates in the 'QE validation' and 'non-QE validation' regions, which have different admixtures of each process. This comparison has an acceptable $\chi^2 = 13.5$ for 9 degrees of freedom. The normalizations are then extrapolated to the signal region, and the excess of events is attributed

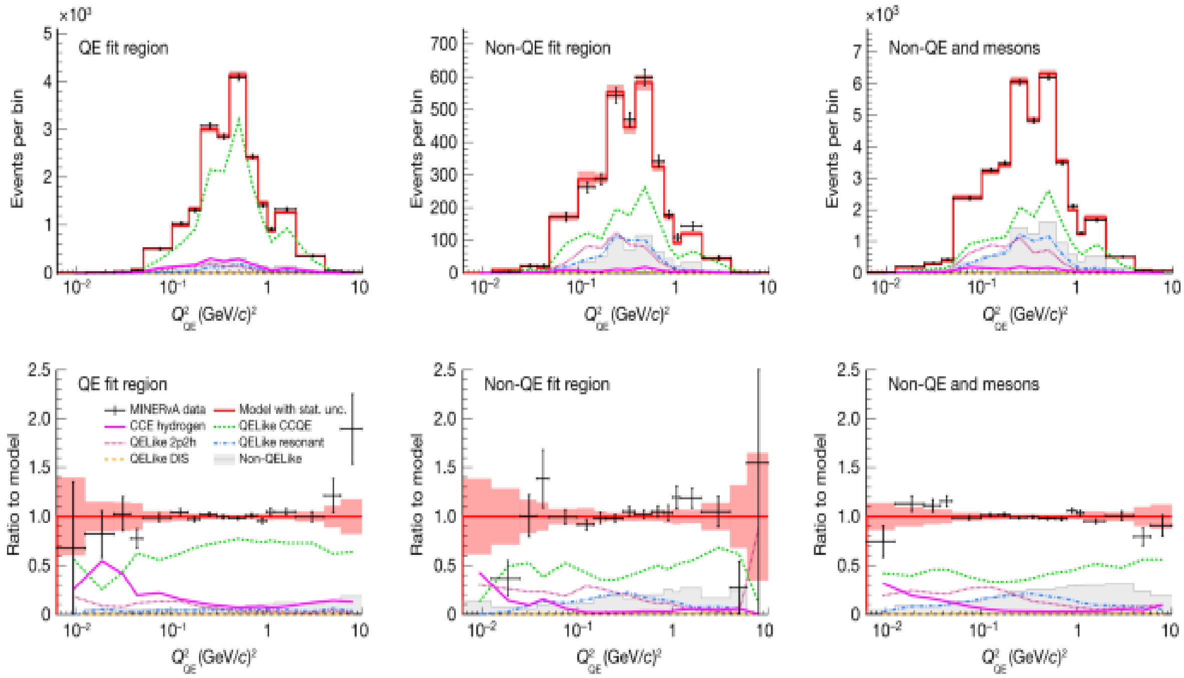
to CCE scattering. By using the data to normalize the background, any modifications in the CCQE cross-section due to changes in F_A from the default model are taken into account. A cross-check of the fit strategy was performed on a proton sample in the beam of neutrinos rather than antineutrinos. Reactions of neutrinos are free of CCE events. The strategy yields good agreement at $Q^2 > 0.2$ GeV/c^2, the threshold above which CCQE proton tracks can be reconstructed by MINERvA. The antineutrino sample, on the other hand, reaches lower Q^2 because the energy deposited by neutron secondary interactions need only span two planes. Figure 2 shows the antineutrino data with the signal and the predicted background levels after the fit, and the ratio to the post-fit model, as a function of Q^2 computed under the hydrogen hypothesis (\({Q}_{{\rm{QE}}}^{2}\)) in the signal region. Figure 3 shows the same distributions in the fit regions. CCQE is the dominant background in the signal and QE fit regions.

Fig. 2: Fitted event distribution and ratio in the signal region.



Event rate (left) and ratio to the post-fit model (right). The vertical error bars around the data points and the error band around the model prediction account for statistical uncertainty (stat. unc.) based on the standard deviation of a Poisson distribution. Note that the (0,0.0125) (GeV/c)^2 and (6.0,10) (GeV/c)^2 bins are not reported due to low statistics. DIS refers to deep inelastic scattering.

Fig. 3: Fitted event distributions and ratios in the control regions.



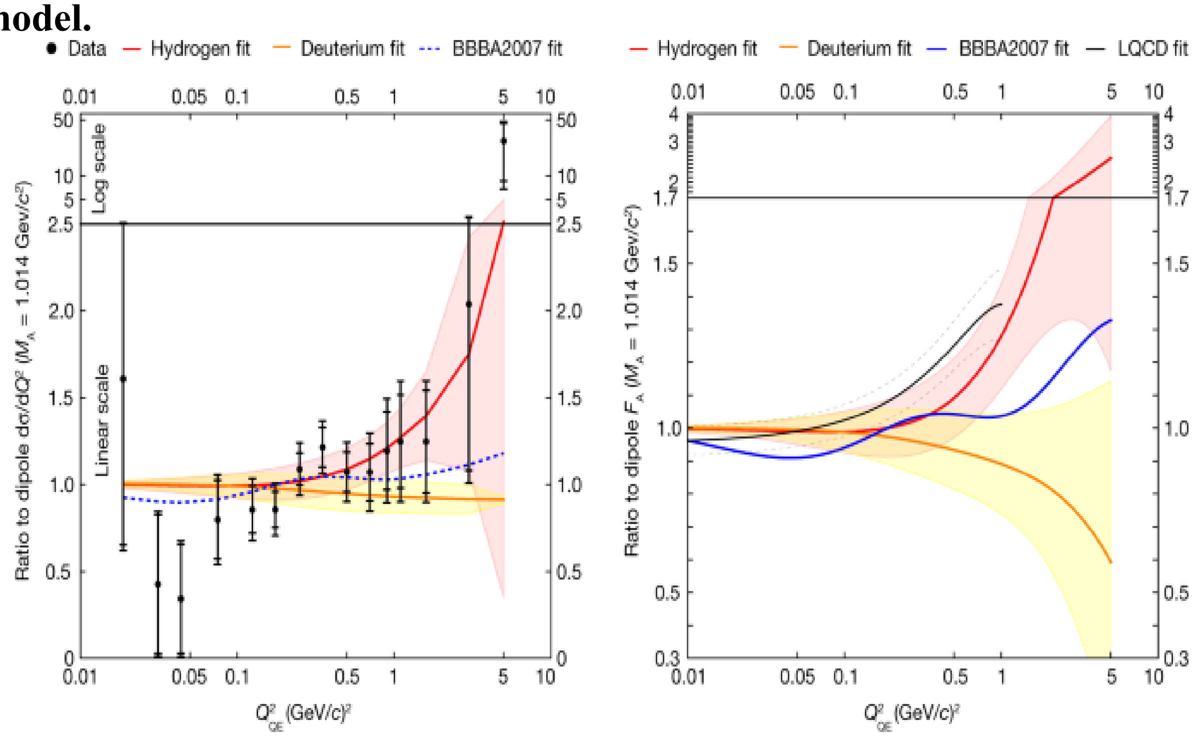
Fitted event distribution in the calculated Q^2 (\({Q}_{{\rm{QE}}}^{2}\)) bins in the QELike sample in the QE, non-QE, as well as non-QE and mesons, regions. Top, event rate per bin; Bottom, ratio to post-fit model. The vertical error bars around the data points and the error band around the model prediction account for 1 standard deviation due to large statistical uncertainties.

The effect of detector resolution on the signal events is corrected using an iterative unfolding method[52,53], and the detector efficiency under the muon phase space cut in each bin is assessed using the simulation. The cross-section is calculated using the fully integrated antineutrino flux[54,55,56] and the total number of hydrogen atoms in the tracker. This measurement does not correct the efficiency loss as a result of the muon phase space cut, which manifests as a restriction on the range of antineutrino energies allowed into the signal sample at each Q^2. Theoretical cross-section predictions need to account for the restricted energy range for comparison with the measurement.

Discussion

This is the first statistically significant measurement, to our knowledge, of the antineutrino CCE scattering on the free proton. We observe 5,580(180) signal events over the estimated background of 12,500 events. The measured cross-section is shown on the left in Fig. 4 in black data points, as a ratio to the cross-section prediction assuming a dipole F_A described below. The analysis is dominated by statistical uncertainties at all Q^2. Systematic uncertainties arise from the small remaining differences, due in part to the regularization, between the post-fit background prediction in each systematic variation of the input model. The dominant systematic uncertainties in this measurement are the neutron secondary interaction in the detector (4.8%), the normalization in the CCQE cross-section (4.5%), the muon energy scale (4.2% from MINOS and 3.1% from MINERvA), the flux (3.9%), neutron FSI (approximately 3%) and the 2p2h process (2.3%).

Fig. 4: Ratios of data and fitted axial vector form factor to a dipole model.



Left, ratios of cross-sections to dipole cross-section with $M_A = 1.014$ GeV/c^2. The inner error bars on the data points account for 1 standard deviation due to statistical uncertainty only, and the full error bars include all sources of systematic uncertainties. Right, ratios to the dipole form factor. The hydrogen (this work) and deuterium[24] F_A fits use the z expansion

formalism; BBBA2007 (ref. [25]) uses a different empirical fit to deuterium and π-electroproduction data; whereas LQCD is a recent fit to lattice QCD calculations[14].

Theory prediction of the measured cross-section requires input from the electromagnetic vector form factors, the axial form factor, the muon momentum and angle restrictions described above, and convolution between the free nucleon cross-section with the antineutrino flux. The electromagnetic form factor used in this study assumes the BBBA2005 (ref. [31]) parameterization. The axial form factor used by most neutrino experiments and generators[36,42,49,57,58] assumes a dipole form, ${F}_{{\rm{A}}}({Q}^{2})={F}_{{\rm{A}}}(0){(1+{Q}^{2}/{M}_{{\rm{A}}}^{2})}^{-2}$, which is an approximation derived from the Fourier transform of an exponential charge distribution. In this ansatz, the shape of F_A depends only on the axial mass term M_A. A more general form, consistent with QCD, is the z expansion formalism[59], which maps the one-dimensional variable $t = -Q^2$ onto a unit circle bounded by ${t}_{{\rm{cut}}}=9{m}_{{\rm{\pi }}}^{2}$, the threshold of three-pion production allowed by the axial current[24]:

$$\begin{array}{rcl}z({Q}^{2},{t}_{{\rm{cut}}},{t}_{0}) & = & \frac{\sqrt{{t}_{{\rm{cut}}}+{Q}^{2}}-\sqrt{{t}_{{\rm{cut}}}-{t}_{0}}}{\sqrt{{t}_{{\rm{cut}}}+{Q}^{2}}+\sqrt{{t}_{{\rm{cut}}}-{t}_{0}}}\\ {F}_{{\rm{A}}}({Q}^{2}) & = & \mathop{\sum }\limits_{k=0}^{{k}_{\max }}{a}_{k}{z}^{k}\end{array}$$
(8)

The hydrogen cross-section is fitted using F_A from the z expansion with $t_0 = -0.75\ (\mathrm{GeV}/c)^2$, ${k}_{\max }=8$. t_0 is chosen so that the Q^2 bins with precise cross-section measurements are distributed symmetrically around $z = 0$. Small variations in t_0 have no impact on the fit result. ${k}_{\max }$ was chosen to be as small as possible while still enabling the fit to describe the data, as tested by a χ^2 statistic. The fit to data includes a bound on the higher order terms[24], such that $|a_k/a_0| \lesssim 5$ and, for $k > 5$, $|a_k/a_0| \lesssim 25/k$. This bound is treated as a Gaussian regularization term during the χ^2 minimization process with a strength parameter λ. The optimal λ of 0.13 was

determined by an L-curve study comparing the minimum χ^2 separated into the comparison to the data and the regularization. The behaviour of F_A at low Q^2 is constrained by $F_\mathrm{A}(0) = -1.2723 \pm 0.0023$, the axial vector coupling as measured in beta decay. A more detailed discussion of the fitting method can be found in the Methods.

The resulting cross-section fit (in red) is shown on the left of Fig. 4 as ratio to a predicted dipole cross-section with $M_A = 1.014\ \mathrm{GeV}/c^2$, together with the predicted cross-section using F_A from the Meyer[24] fit (in yellow) on deuterium data and a fit derived jointly from deuterium and pion electroproduction data (BBBA2007, in dotted blue)[25]. The resulting form factor as a ratio to the dipole form factor is shown on the right. The cross-section ratio scales approximately linearly with F_A ratios due to suppression of the A term in equations (4) and (5). The nucleon axial radius from the fit to this result is ${r}_{{\rm{A}}}\equiv \sqrt{\left\langle {{r}_{{\rm{A}}}}^{2}\right\rangle }=0.73(17)\ {\rm{fm}}$.

This result is the first statistically significant measurement, as far as we are aware, of the axial vector form factor on free protons without nuclear corrections or other theoretical assumptions. Theoretical uncertainties from the carbon background have been minimized by data-driven methods. By providing a precise and reliable prediction for the charged-current elastic scattering from nucleons, neutrino measurements on higher Z nuclei can benefit from better constrained nucleon effects to expose the nuclear effects. The method developed in this study will enable future experiments with hydrogen content in the target[18,19] to make further measurements of the axial form factor. Future experiments with intrinsic three-dimensional capability would be able to observe the directions of low-energy neutron candidates, and improve the low Q^2 measurement with more statistics.

Methods

Terminology of the signal process

This article refers to the $\bar{\nu }+H\to {\mu }^{+}n$ scattering process as the CCE process to distinguish it from the CCQE process that

experiments have measured in scattering from bound nucleons in nuclei. The signal process is an isolated, two-body reaction that conserves the invariant mass, and is ‘elastic’ in the sense that all energy from the neutrino and the target proton is transferred to the final-state muon and neutron.

Detector and neutrino beam

The MINERvA[11] detector, shown in Extended Data Fig. 1, is located just upstream of the MINOS ND[36] in the NuMI beamline[34] at Fermilab. The main injector supplies a beam of 120 GeV protons striking a graphite target 1.04 km upstream of the detector. Secondary mesons produced at the target are charge selected by two magnetic horns to enhance the ν_μ production in the forward horn current mode and the \({\bar{\nu }}_{\mu }\) production in the reverse horn current (RHC) mode. The resulting neutrino or antineutrino beam travels at a downward angle of 0.059 radian and has a broad energy profile. In particular, the antineutrino beam in the RHC mode peaks at 5.5 GeV (ref. [60]). Extended Data Fig. 8 (right) shows the flux distribution.

Extended Data Fig. 2 describes the detector components. The MINERvA detector is capable of measuring \(\bar{\nu }p\to {\mu }^{+}n\) events over a large range of Q^2, up to 7 $(\text{GeV}/c)^2$ by tracking the muon through the scintillator target and reconstructing the secondary interaction of the produced neutrons. Each plane in the active tracker region consists of scintillator strips with a triangular 33 mm (transverse) by 17 mm (along beam) cross-section interlocked together. Particles typically deposit energy in multiple adjacent strips, which gives position resolution smaller than the size of the strips. To be reconstructed for this analysis, muons produced in the scintillator target must exit from the back of the MINERvA detector and be measured by the MINOS ND[36]. The magnetized steel in MINOS provides an average 1.3 T field, and the interspersed scintillator planes measure the length and curvature of the muon tracks to determine the muon’s charge and momentum as it enters the MINOS ND. Muon tracks in MINERvA are reconstructed and fitted with a Kalman filter resulting in 3.1 mm tracking resolution. The total muon energy and angle resolutions are 6% and 0.06°, respectively[61]. Muons well tracked by the MINOS ND usually

have accepted angles to the beam $\theta_\mu < 20°$ and momenta in the range 1.5 GeV/$c < p_\mu < 20$ GeV/c, respectively.

Signal definition

The CCE, $\bar{\nu}H\to {\mu}^{+}n$, signal is a subset of the quasi-elastic-like (QELike) sample in antineutrino scattering events. Precisely, 'QELike' events are those with a μ^+ and only nucleons and remnant nuclei in the final state. In addition, events with energetic protons that produce reconstructable tracks, corresponding to a kinetic energy threshold of about 120 MeV, are excluded from the 'QElike' category in the antineutrino sample. The remaining charged-current antineutrino events, with other final-state particles or with sufficiently energetic protons, are labelled as 'non-QELike'. The QELike category contains, apart from the signal events, CCQE interactions on protons bound in carbon nuclei, interactions in which mesons are produced but later absorbed in the nucleus, and multinucleon knockout events (2p2h), in which two target nucleons participate in the scattering process. The antineutrino QELike category is naturally neutron-rich because of the net negative charge transferred to the target. Neutron detection is inefficient and does not give a strong constraint on neutron energy; therefore, it is not possible to exclude classes of events with multiple neutrons by counting detectable neutrons or measuring their energies. Signal events are instead constrained by the consistency of the direction of the neutron with the kinematics of the charged-current elastic scattering signal process.

Reference interaction and detector model

A complete Monte Carlo simulation of the MINERvA detector, data acquisition and event reconstruction chain was developed to realistically predict the detector efficiency and resolution. The detector was simulated with the GEANT4 toolkit[41]. Signal and background events are generated by the GENIE[42] v.2.12.6 neutrino event generator with data-driven modifications. GENIE describes the carbon nucleus with a relativistic Fermi gas (RFG) model that assumes the nucleons are free, with the Fermi momentum maximum of $K_F = 221$ MeV. The single nucleon momentum is

extended beyond the Fermi momentum cutoff to account for short-range correlations[62]. For the CCQE reactions, the RFG model is reweighted to the prediction of a spectral function model[50,63] used in NuWro[49], which incorporates both the shell structure of the nucleus and the nucleon correlation. Without the spectral function features, the initial state model is not sufficiently consistent with the carbon CCQE background to provide an accurate constraint. The reference CCE and CCQE cross-sections are modelled with a dipole axial form factor with $M_A = 0.99\ \mathrm{GeV}/c^2$ and the BBBA2005 (ref. [31]) vector form factors. The Valencia 2p2h model[44] used to describe the knockout of multiple nucleons is tuned by adding a large enhancement to match other MINERvA measurements[43]. The FSI is simulated by the GENIE INTRANUKE-hA[42] model. The model emulates a full-scale intranucleus cascade[64] in a way that is easily varied for systematic uncertainty studies. Some of the FSI reactions that change the final-state particle content in the detector are the inelastic FSI, pion production, pion absorption and pion and nucleon charge exchange. The elastic component of the FSI simulation in the CCQE channel is replaced with the no-FSI component through reweighting to circumvent an error in the INTRANUKE-hA kinematic implementation[65]. Non-CCQE interactions, such as resonant pion production, are based on the Rein–Seghal[66] model, and the deep inelastic scattering is based on the Bodek–Yang[67] model.

Neutron interaction and reconstruction

Neutrons produced in neutrino interactions can interact in the detector through elastic or inelastic interactions on hydrogen or carbon. Neutron capture rates are negligible at MeV energies[68]. Elastic scattering on hydrogen $n^1\mathrm{H} \rightarrow np$ or inelastic scattering on carbon, such as $n^{12}\mathrm{C} \rightarrow np^{11}\mathrm{B}$, can produce protons with enough energy to be observed in the detector. Neutron interactions are modelled using GEANT4. However, GEANT4 predictions of some inelastic neutron-C cross-sections are not in agreement with measurements of neutron scattering on scintillators at energies similar to our signal reaction[69]; therefore, we reweight the probability of neutron scattering from the GEANT4 prediction to be consistent with the recent models[70] that do agree with these data.

The MINERvA experiment has developed algorithms to reconstruct energy deposits from neutrons that produce ionizing particles by scattering off the material in the detector[37]. The daughter particles, primarily protons, are usually observed as isolated hits in the detector, far from the interaction vertex. If the particle is energetic enough to pass through two or more planes, a three-dimensional position of the interaction can be reconstructed and the direction from the vertex measured. The angle between the reconstructed and true neutron directions was assessed with the detector simulation. The distribution can be modelled as the sum of two exponential distributions $N(\theta)={A}_{1}/{\Theta }_{1}\exp (-\theta /{\Theta }_{1})+{A}_{2}/{\Theta }_{2}\exp (-\theta /{\Theta }_{2})$, where the values of A_1 and A_2 are such that 40% of the neutrons have a narrow $\Theta_1 = 6°$ exponential slope and the remainder have a $\Theta_2 = 20°$ exponential slope. The heavy tail of this distribution probably results from the undetected scatters of neutrons in the detector before a detected scatter is observed. The intrinsic neutron angular resolution is convoluted with physics-driven neutron direction to enable constraint on the carbon CCQE events.

Extended Data Fig. 2 illustrates the schematics of the angular variables $\delta\theta_R$ and $\delta\theta_P$, and shows raw data and fitted model distributions. The $\delta\theta_P$ distribution is symmetric about $\delta\theta_P = 0$, largely due to Fermi momentum in the CCQE background, whereas for the CCE signal the smearing is dominated by secondary interaction effects. There are bulk shifts in the peak positions of $\delta\theta_R$ distribution from the non-CCQE components due to the neutron assumption required in the direction prediction. Whereas the transverse component of the predicted neutron momentum is fixed by the muon's transverse momentum (equation (3)), the component parallel to the neutrino beam depends on the invariant hadron mass of the assumed final state (${M}_{{\rm{n}}}^{2}$ in equation (2)), resulting in a systematic shift when the final-state hadronic system is not a single neutron. Extended Data Fig. 3 shows the Monte Carlo event rates for a few interaction types at different Q^2.

Event selection

Events with a muon track starting in the scintillator tracker region and ending in the MINOS ND with a positive charge identification are selected for analysis. Events with additional tracks formed at the vertex are rejected because they indicate the presence of charged mesons or energetic protons. The recoil energy is defined as the total energy outside a 100 mm sphere from the vertex and not associated with the muon track. A Q^2-dependent maximum recoil energy cut was applied to reduce the fraction of non-QELike background (Extended Data Table 1).

As described above, neutrons become detectable when they inelastically interact in the detector to create secondary particles. Neutron candidates in the detector are reconstructed by an algorithm that combines detector hits not attached to an interaction vertex and known particle tracks. Neutron candidates whose hits span two views or more produce three-dimensional position information, and only events with such candidates are selected for this analysis. Even at this early stage, the sample is already dominated by events with a final-state neutron from an antineutrino interaction.

In the CCE \(\bar{\nu }{\rm{H}}\) reaction, the energy transfer q_0 from the neutrino to the initial state proton is proportional to the Q^2:

$${q}_{0}=\frac{{Q}^{2}}{2{M}_{{\rm{n}}}}.$$
(9)

Leading neutron candidates with detectable energy above this physics limit cannot originate from the signal reaction, so such events are rejected. The QELike selection also restricts the number of isolated energy deposits outside the vertex region. Events with a single deposit whose energy resembles proton energy deposits with at least 10 MeV per strip are defined to be ‘QELike’. Because of the inefficient neutron selection, events with multiple energy deposits are predominately events with particles other than neutrons, such as photons from π^0 and mis-reconstructed $\pi^{\pm}$. Events with multiple candidates are defined to be in the ‘non-QELike’ sideband sample.

Control regions are formulated on the basis of opening angular separations between the reconstructed neutron candidate and the expected neutron direction under the hydrogen hypothesis. The angular variables are defined according to equation (7), with regions formulated with varying fractions of

QELike backgrounds, as shown in Fig. 1. The signal selection occupies the central $-10° < \delta\theta_{P}/\delta\theta_{R} < 10°$ region, whereas the 'non-QE and mesons' region encompasses all angular space outside the named regions. A bin-by-bin background constraint is performed for both the QELike and non-QELike events in bins of ${Q}_{{\rm{QE}}}^{2}$, which is the Q^2 computed under the hydrogen hypothesis. Equation (10) shows the χ^2 function minimized in the fit:

$${\chi }^{2}=\sum _{S,i}\frac{{\left(\left[{\sum }_{C}{w}_{C,i}{N}_{C,S,i}^{{\rm{mc}}}\right]-{N}_{S,i}^{{\rm{data}}}\right)}^{2}}{{N}_{S,i}^{{\rm{data}}}}+{\lambda }_{S}\sum _{C}\mathop{\sum }\limits_{j=1}^{N-2}{({w}_{C,j}+{w}_{C,j+2}-2{w}_{C,j+1})}^{2}. \quad (10)$$

where ${N}_{C,S,i}^{{\rm{mc}}}$ is the Monte Carlo event rate in the ith ${Q}_{{\rm{QE}}}^{2}$ bin in the interaction type category C, and in the angular control sample S. ${N}_{S,i}^{{\rm{data}}}$ is the data rate in the ith ${Q}_{{\rm{QE}}}^{2}$ bin in a control sample S. Each category used in the fit receives a weight $w_{C,i}$; categories not fitted receive a constant weight of 1. A regularization term is added to the χ^2 function to ensure the weights across the bins change smoothly. λ_S is the regularization strength and is a meta parameter tuned to ensure consistency between the fit and validation regions. To do so, the fit is performed on the central value model for a range of λ_S. An average χ^2 for the validation regions, weighted by their event rates, is then calculated. The χ^2 from the QE validation region dominates this calculation due to its high statistics. The value of λ_S is chosen through an L-curve study. A separate fit using these parameters is done for every source of systematic uncertainty to evaluate its effect on the fit.

The QELike categories constrained are CCQE on carbon, 2p2h and resonant events whose pions are absorbed through FSI. The CCE signal category is not tuned because its contribution is small in the control regions. QE carbon is the primary background type in the QELike sample. Events with mesons in the final states, such as single π^0, $\pi^{\pm}$ and multiple pions, are primarily constrained with the non-QELike sample. The control regions used for fitting are the 'QE', 'non-QE' and 'non-QE and mesons', in the QELike

sample, and all regions combined in the non-QELike sample. The QE and non-QE validation regions in the QELike sample are used to assess the goodness-of-fit. The results of the fit and the ratio to the fitted models in the validation regions are shown in Extended Data Fig. 4. The total event rate in the non-QE validation region is about one-fifth of the QE validation region.

Validation of CCQE model and background subtraction technique using ν_μ events

The validity of the background constraint techniques with the above model was tested using a neutrino CCQE sample with no CCE scattering from the free hydrogen. Events with a negatively charged muon matched to MINOS and a reconstructed proton track are selected. As with the CCE reconstruction, we compute the theoretical nucleon direction, this time a final-state proton, on an event-by-event basis and compare it with the proton track direction to classify the event according to the same angular region scheme. The track reconstruction efficiency at MINERvA for protons scales with their momenta, and below 450 MeV/c (100 MeV in kinetic energy) the protons are no longer reconstructed. Therefore, QE protons with $Q^2 < 0.2$ (GeV/c)^2 are very rare, with events dominated by 2p2h background events whose reconstructed Q^2 based on the QE hypothesis are small.

A fit was performed using the control samples. Extended Data Fig. 5 shows both the fitted neutrino event ratio to the simulation in the signal and QE regions analogous to Fig. 2 and the $\delta\theta_{\mathrm{R}}$ in this sample. The fitted Monte Carlo adequately describes data when proton tracks from CCQE are reliably reconstructed. The small discrepancy at Q_{QE}^{2} between 0.2 (GeV/c)^2 and 0.4 (GeV/c)^2 is covered by a 100% uncertainty in the 2p2h component in that restricted region, as shown by the $\delta\theta_{\mathrm{R}}$ plot at the bottom of the figure, and is the probable source of the discrepancy, given that the 2p2h component only begins to contribute significantly in this Q_{QE}^{2} range. Such a systematic deviation in the carbon model would be well within the systematic uncertainties applied to our CCE background analysis in the antineutrino beam.

Unfolding

The signal event rate obtained after background subtraction contains smearing effects from detector resolution. This analysis, similar to many recent results from the MINERvA Collaboration, undoes the smearing by means of an iterative unfolding algorithm developed by D'Agostini[52]. The inputs to the algorithm are the smearing matrix obtained through the Monte Carlo study of the true and reconstructed Q^2 in the signal sample and the number of iteration steps, N_i, as a regularization parameter. When N_i is too small, the unfolded distribution will not have converged to the true kinematic distribution, but higher values of N_i increase statistical variances of the reconstructed distribution. To find the value of N_i that balances model bias and statistical uncertainties, we form statistical universes of a toy model obtained by reweighting the Monte Carlo signal sample. The toy model differs from the CCE model by about 10% at lower Q^2, and 20% at higher Q^2, to mimic the shape of the discrepancy between data and the model. An ensemble of pseudo-experiments is created by changing the value in each bin of the toy model to a random number generated by a Poisson distribution, with the mean set at the central value of the toy model in that bin. A total of 1,000 statistical universes have been made and $N_i = 4$ minimizes the median and mean of the χ^2 of the result of the pseudo-experiments compared to the hypothesis of the toy model.

The unfolded data are corrected for the predicted efficiency obtained from the Monte Carlo study. The efficiency, shown in Extended Data Fig. 6, increases with higher Q^2 until 0.5 (GeV/c)^2 as the rate of inelastic interactions that produce protons over the reconstruction threshold increases. For Q^2 greater than 0.5 (GeV/c)^2, however, the signal efficiency drops due to two factors. First, the neutron angle with respect to the neutrino direction increases, and therefore the neutron passes through less detector material on average before exiting the detector. Second, the muon is produced at a wider angle with respect to the neutrino and therefore is less likely to be reconstructed in the downstream MINOS ND.

Uncertainties

Extended Data Fig. 7 shows the breakdown of the fractional systematic and statistical uncertainties in our measurement. We assess systematic

uncertainties in six broad categories. Modelling uncertainties in GENIE are broken into the FSI and cross-section model uncertainties. The muon reconstruction uncertainties account for resolutions in the muon energy, direction and the reconstruction efficiency. The low recoil fit category is due to uncertainties in the tuned 2p2h model. The 'other' category includes target mass uncertainties and a set of particle response uncertainties. The largest particle response uncertainties are related to the neutron interactions in the detector. The 'neutron reweight' uncertainty is due to the neutron elastic and inelastic cross-sections assumed in the detector; it is assessed for each neutron and assigned a larger value for lower kinetic energy. The neutron interaction uncertainty accounts for the discrepancy in the energy deposited in the neutron candidates we observed in a previous study[37]. These are the largest uncertainties in the 'other' category, and they mainly affect the low Q^2 cross-section. This analysis is dominated by the statistical uncertainty, in part because of the large background subtraction.

Cross-section calculation and the z expansion fit of $F_A(Q^2)$

The cross-section measured in this analysis is the flux-integrated cross-section for the geometry of the MINERvA detector. The measured event and predicted background rates are shown in Supplementary Table 1. The efficiency corrected event rates are divided by the total number of hydrogen atoms and protons on target (Supplementary Table 2). The measured cross-section is shown in tabulated form in Supplementary Table 3, and the total covariance and statistical-only covariance matrices of the measurement are shown in Supplementary Tables 6 and 7, respectively.

Calculating the integrated cross-section in each bin needs to account for the muon phase space cuts described above. The effect of the phase space restriction is manifested as a restricted range of neutrino energy available in the selected sample at each Q^2 point, ultimately reducing the differential cross-section at each Q^2 bin because the acceptance-corrected event rate is divided by the fully integrated neutrino flux for this measurement. Extended Data Fig. 8 (right) illustrates the accepted neutrino energy at each Q^2. Therefore, obtaining a fit of F_A requires the convolution of equation (4) with the antineutrino flux from the RHC configuration of the NuMI[34] neutrino

beam with the Q^2-dependent energy cutoff. The vector form factors used in this analysis are parameterized from electron scattering data[31] used commonly by neutrino Monte Carlo generators[42,49,58].

The axial vector form factor is also fitted using the z expansion[59] formalism. We adopt the procedure of a fit to deuterium data by Meyer et al.[24]. The z expansion for $F_{\rm A}$ is a polynomial of z with coefficients a_k reproduced here for convenience.

$${F}_{{\rm{A}}}({Q}^{2})=\mathop{\sum }\limits_{k=0}^{{k}_{\max }}{a}_{k}{z}^{k}$$
(11)
$$z=\frac{\sqrt{{t}_{{\rm{cut}}}+{Q}^{2}}-\sqrt{{t}_{{\rm{cut}}}-{t}_{0}}}{\sqrt{{t}_{{\rm{cut}}}+{Q}^{2}}+\sqrt{{t}_{{\rm{cut}}}-{t}_{0}}}$$
(12)
$$\mathop{\sum }\limits_{k=n}^{\infty }k(k-1)\ldots (k-n+1){a}_{k}=0,n\in (0,1,2,3)$$
(13)

Values of the coefficients a_k are constrained by $F_{\rm A}(0) = -1.2723$, and a set of four coefficient sum rules (equation (13)) determined by the high Q^2 behaviour required by QCD[71,72]. For a given value of ${k}_{\max }$, ${N}_{a}={k}_{\max }-4$ coefficients are free parameters: $({a}_{1},...,{a}_{{k}_{\max }-4})$. The error function involved in the fit is a χ^2 function with a regularization term of the form $\lambda \left[{\sum }_{k=1}^{5}{\left(\frac{{a}_{k}}{5{a}_{0}}\right)}^{2}+{\sum }_{k=5}^{{k}_{\max }}{\left(\frac{k{a}_{k}}{25{a}_{0}}\right)}^{2}\right]$ based on the expected fall off of $a_k \propto k^4$ at large k (ref. [24]). We choose $t_0 = -0.75\ ({\rm GeV}/c)^2$ in our analysis to centre our data points around $z = 0$. Small variations in t_0 have no effect on the final fit results. ${k}_{\max }$ and λ are parameters that affect the functional form of $F_{\rm A}$ and the uncertainties of the form factor. Larger ${k}_{\max }$ increases allow for a more ‘curvy’ $F_{\rm A}$ that might overfit and follow statistical variations of the data, whereas a non-zero λ suppresses this behaviour for larger ${k}_{\max }$. At ${k}_{\max }=6$, the value of λ

has little effect on the low Q^2 behaviour of the fit but prevents F_A from adequately describing the data at larger Q^2. The fit at ${k}_{\max }=8$ enables higher Q^2 fit to follow the data points, but also exposes the low-Q^2 region to a small dependence on λ, which manifests as a few per cent variation in the calculation of nucleon axial radius measured on a proton, sufficiently covered by the uncertainties from the fit. The behaviour of the regularization term and the data χ^2 term were studied and the fit at $\lambda = 0.13$ was chosen according to the maximum curvature, or L-curve, criterion[73]. Extended Data Fig. 8 (left) shows the final fit to data. In general, varying the z expansion parameters, such as ${k}_{\max }$, t_0 and λ, results in changes below 10% of the total uncertainty. Supplementary Tables 4 and 5 show the fit result and the correlation matrix using ${k}_{\max }=8,\lambda =0.13$ and another fit with ${k}_{\max }=6,\lambda =0$.

Data availability

The extracted cross-section data reported in this study are available in the extended figures and tables section, and will be made available for public access on https://minerva.fnal.gov/data-release-page/.

Code availability

The MINERvA Collaboration develops and maintains the code used for the simulation of the experimental apparatus and statistical analysis of the raw data used in this result. This code is shared among the collaboration, but not publicly distributed. Enquiries regarding the algorithms and methods used in this result may be directed to the corresponding author.

References

1. Chambers, E. E. & Hofstadter, R. Structure of the proton. *Phys. Rev.* **103**, 1454–1463 (1956).

2. Ernst, F. J., Sachs, R. G. & Wali, K. C. Electromagnetic form factors of the nucleon. *Phys. Rev.* **119**, 1105–1114 (1960).

3. Mann, W. A. et al. Study of the reaction $\nu + n \rightarrow \mu^- + p$. *Phys. Rev. Lett.* **31**, 844–847 (1973).

4. Barish, S. J. et al. Study of neutrino interactions in hydrogen and deuterium. Description of the experiment and study of the reaction neutrino $\nu + d \rightarrow \mu^- + p + p_s$. *Phys. Rev. D* **16**, 3103 (1977).

5. Baker, N. J. et al. Quasielastic neutrino scattering: a measurement of the weak nucleon axial vector form-factor. *Phys. Rev. D* **23**, 2499–2505 (1981).

6. Miller, K. L. et al. Study of the reaction $\nu_\mu d \rightarrow \mu^- p p_s$. *Phys. Rev. D* **26**, 537–542 (1982).

7. Kitagaki, T. et al. High-energy quasielastic $\nu_\mu n \rightarrow \mu^- p$ scattering in deuterium. *Phys. Rev. D* **28**, 436–442 (1983).

8. Allasia, D. et al. Investigation of exclusive channels in neutrino/anti-neutrino deuteron charged current interactions. *Nucl. Phys. B* **343**, 285–309 (1990).

9. Kitagaki, T. et al. Study of $\nu d \rightarrow \mu^- p p_s$ and $\nu d \rightarrow \mu^- \Delta^{++}(1232) n_s$ using the BNL 7-foot deuterium filled bubble chamber. *Phys. Rev. D* **42**, 1331–1338 (1990).

10. Singh, S. K. The effect of final state interactions and deuteron binding in $\nu d \rightarrow \mu^- p p$. *Nucl. Phys. B* **36**, 419–435 (1972).

11. Aliaga, L. et al. Design, calibration, and performance of the MINER*ν*A detector. *Nucl. Instrum. Meth. Phys. Res. A* **743**, 130–159 (2014).

12. Kronfeld, A. S. et al. Lattice QCD and neutrino-nucleus scattering. *Eur. Phys. J. A* **55**, 196 (2019).

13. Alexandrou, C. et al. Nucleon axial and pseudoscalar form factors from lattice QCD at the physical point. *Phys. Rev. D* **103**, 034509 (2021).

14. Lin, H.-W. Nucleon helicity generalized parton distribution at physical pion mass from lattice QCD. *Phys. Lett. B* **824**, 136821 (2022).

15. Meyer, A. S., Walker-Loud, A. & Wilkinson, C. Status of lattice QCD determination of nucleon form factors and their relevance for the few-GeV neutrino program. *Annu. Rev. Nucl. Part. Sci.* **72**, 205–232 (2022).

16. Ayres, D. S. et al. *The NOvA Technical Design Report* https://doi.org/10.2172/935497 (2007).

17. Abe, K. et al. The T2K experiment. *Nucl. Instrum. Meth. Phys. Res. A* **659**, 106–135 (2011).

18. Abi, B. et al. Deep Underground Neutrino Experiment (DUNE) Far Detector Technical Design Report, Volume II: DUNE Physics. Preprint at https://arxiv.org/abs/2002.03005 (2020)

19. H.-K. Proto-Collaboration et al. Hyper-Kamiokande Design Report. Preprint at https://arxiv.org/abs/1805.04163 (2018).

20. An, F. et al. Neutrino physics with JUNO. *J. Phys. G* **43**, 030401 (2016).

21. Miller, G. A. Defining the proton radius: a unified treatment. *Phys. Rev. C* **99**, 035202 (2019).

22. Xiong, W. et al. A small proton charge radius from an electron–proton scattering experiment. *Nature* **575**, 147–150 (2019).

23. Atac, H., Constantinou, M., Meziani, Z.-E., Paolone, M. & Sparveris, N. Measurement of the neutron charge radius and the role of its constituents. *Nat. Commun.* **12**, 1759 (2021).

24. Meyer, A. S., Betancourt, M., Gran, R. & Hill, R. J. Deuterium target data for precision neutrino-nucleus cross sections. *Phys. Rev. D* **93**, 113015 (2016).

25. Bodek, A., Avvakumov, S., Bradford, R. & Budd, H. S. Vector and axial nucleon form factors: a duality constrained parameterization. *Eur.*

Phys. J. C **53**, 349–354 (2008).

26. Megias, G. D., Bolognesi, S., Barbaro, M. B. & Tomasi-Gustafsson, E. New evaluation of the axial nucleon form factor from electron- and neutrino-scattering data and impact on neutrino-nucleus cross sections. *Phys. Rev. C* **101**, 025501 (2020).

27. Bernard, V., Elouadrhiri, L. & Meißner, U.-G. Axial structure of the nucleon. *J. Phys. G* **28**, R1 (2001).

28. Andreev, V. A. et al. Measurement of muon capture on the proton to 1% precision and determination of the pseudoscalar coupling g_P. *Phys. Rev. Lett.* **110**, 012504 (2013).

29. Fanourakis, G. et al. Study of low-energy anti-neutrino interactions on protons. *Phys. Rev. D* **21**, 562 (1980).

30. Llewellyn Smith, C. H. Neutrino reactions at accelerator energies. *Phys. Rep.* **3**, 261–379 (1972).

31. Bradford, R., Bodek, A., Budd, H. S. & Arrington, J. A new parameterization of the nucleon elastic form factors. *Nucl. Phys. B Proc. Suppl.* **159**, 127–132 (2006).

32. Borah, K., Hill, R. J., Lee, G. & Tomalak, O. Parametrization and applications of the low-Q^2 nucleon vector form factors. *Phys. Rev. D* **102**, 074012 (2020).

33. Goldberger, M. L. & Treiman, S. B. Decay of the pi meson. *Phys. Rev.* **110**, 1178–1184 (1958).

34. Adamson, P. et al. The NuMI neutrino beam. *Nucl. Instrum. Meth. Phys. Res. A* **806**, 279–306 (2016).

35. Olive, K. A. Review of particle physics. *Chin. Phys. C* **40**, 100001 (2016).

36. Michael, D. G. et al. The magnetized steel and scintillator calorimeters of the MINOS experiment. *Nucl. Instrum. Meth. Phys. Res. A* **596**, 190–

228 (2008).

37. Elkins, M. et al. Neutron measurements from antineutrino hydrocarbon reactions. *Phys. Rev. D* **100**, 052002 (2019).

38. Schiavilla, R., Benhar, O., Kievsky, A., Marcucci, L. E. & Viviani, M. Polarization transfer in \({}^{4}{\rm{H}}{\rm{e}}{(\overrightarrow{e},{e}^{{\prime} }\overrightarrow{p})}^{3}{\rm{H}}\): is the ratio G_{Ep}/G_{Mp} modified in the nuclear medium? *Phys. Rev. Lett.* **94**, 072303 (2005).

39. Strauch, S. et al. Polarization transfer in the \({}^{4}{\rm{H}}{\rm{e}}{(\overrightarrow{e},{e}^{{\prime} }\overrightarrow{p})}^{3}{\rm{H}}\) reaction up to $Q^2 = 2.6(\mathrm{GeV}/c)^2$. *Phys. Rev. Lett.* **91**, 052301 (2003).

40. Thomas, A. W. *Role of Quarks in Nuclear Structure* (Oxford Univ. Press, 2020); https://oxfordre.com/physics/view/10.1093/acrefore/9780190871994.001.0001/acrefore-9780190871994-e-1.

41. Agostinelli, S. et al. Geant4–a simulation toolkit. *Nucl. Instrum. Meth. Phys. Res. A* **506**, 250–303 (2003).

42. Andreopoulos, C. The GENIE neutrino Monte Carlo generator. *Acta Phys. Polon. B* **40**, 2461–2475 (2009).

43. Rodrigues, P. A., Demgen, J. & Miltenberger, E. et al. Identification of nuclear effects in neutrino-carbon interactions at low three-momentum transfer. *Phys. Rev. Lett.* **116**, 071802 (2016).

44. Nieves, J., Ruiz Simo, I. & Vicente Vacas, M. J. Inclusive charged-current neutrino-nucleus reactions. *Phys. Rev. C* **83**, 045501 (2011).

45. Wilkinson, C., Rodrigues, P., Cartwright, S., Thompson, L. & McFarland, K. Reanalysis of bubble chamber measurements of muon-neutrino induced single pion production. *Phys. Rev. D* **90**, 112017 (2014).

46. Rodrigues, P., Wilkinson, C. & McFarland, K. Constraining the GENIE model of neutrino-induced single pion production using reanalyzed bubble chamber data. *Eur. Phys. J. C* **76**, 474 (2016).

47. Nieves, J., Amaro, J. E. & Valverde, M. Inclusive quasi-elastic neutrino reactions. *Phys. Rev. C* **70**, 055503 (2004).

48. Stowell, P. et al. Tuning the GENIE pion production model with MINERvA data. *Phy. Rev. D* **100**, 072005 (2019).

49. Juszczak, C., Nowak, J. A. & Sobczyk, J. T. Simulations from a new neutrino event generator. *Nucl. Phys. B* **159**, 211–216 (2006).

50. Ankowski, A. M. & Sobczyk, J. T. Construction of spectral functions for medium-mass nuclei. *Phys. Rev. C* **77**, 044311 (2008).

51. Benhar, O., Farina, N., Nakamura, H., Sakuda, M. & Seki, R. Electron- and neutrino-nucleus scattering in the impulse approximation regime. *Phys. Rev. D* **72**, 053005 (2005).

52. D'Agostini, G. A multidimensional unfolding method based on Bayes' theorem. *Nucl. Instrum. Meth. Phys. Res. A* **362**, 487–498 (1995).

53. D'Agostini, G. Improved iterative Bayesian unfolding. Preprint at https://doi.org/10.48550/arXiv.1010.0632 (2010).

54. Valencia, E. et al. Constraint of the MINERνA medium energy neutrino flux using neutrino-electron elastic scattering. *Phys. Rev. D* **100**, 092001 (2019).

55. Ruterbories, D. et al. Constraining the NuMI neutrino flux using inverse muon decay reactions in MINERvA. *Phys. Rev. D* **104**, 092010 (2021).

56. Zazueta, L. et al. Improved constraint on the MINERvA medium energy neutrino flux using \(\bar{\nu } {e}^{-}\to \bar{\nu } {e}^{-}\) data. Preprint at https://doi.org/10.48550/arXiv.2209.05540 (2022).

57. Abe, K. et al. Constraint on the matter–antimatter symmetry-violating phase in neutrino oscillations. *Nature* **580**, 339–344 (2020).

58. Hayato, Y. A neutrino interaction simulation program library NEUT. *Acta Phys. Polon. B* **40**, 2477–2489 (2009).

59. Hill, R.J. & Paz, G. Model-independent extraction of the proton charge radius from electron scattering. *Phys. Rev. D* https://doi.org/10.1103/physrevd.82.113005 (2010).

60. Aliaga, L. et al. Neutrino flux predictions for the NuMI beam. *Phys. Rev. D* **94**, 092005 (2016).

61. Bashyal, A. et al. Use of neutrino scattering events with low hadronic recoil to inform neutrino flux and detector energy scale. *J. Instrum.* **16**, P08068 (2021).

62. Bodek, A. & Ritchie, J. L. Fermi-motion effects in deep-inelastic lepton scattering from nuclear targets. *Phys. Rev. D* **23**, 1070–1091 (1981).

63. Benhar, O., Day, D. & Sick, I. An archive for quasi-elastic electron-nucleus scattering data. Preprint at https://arxiv.org/abs/nucl-ex/0603032 (2006).

64. Dytman, S. Final state interaction models in neutrino nucleus cross sections. *AIP Conf. Proc.* **1189**, 51–59 (2009).

65. Harewood, L. & Gran, R. Elastic hadron-nucleus scattering in neutrino-nucleus reactions and transverse kinematics measurements. Preprint at https://doi.org/10.48550/arXiv.1906.10576 (2019).

66. Rein, D. & Sehgal, L. M. Neutrino-excitation of baryon resonances and single pion production. *Ann. Phys.* **133**, 79–153 (1981).

67. Bodek, A. & Yang, U. K. Modeling deep inelastic cross-sections in the few GeV region. *Nucl. Phys. Proc. Suppl.* **112**, 70–76 (2002).

68. Brown, D. A. et al. ENDF/B-VIII.0: the 8th major release of the nuclear reaction data library with CIELO-project cross sections, new

standards and thermal scattering data. *Nucl. Data Sheets* **148**, 1–142 (2018) .

69. Del Guerra, A. A compilation of n-p and n-C cross sections and their use in a Monte Carlo program to calculate the neutron detection efficiency in plastic scintillator in the energy range 1–300 MeV. *Nucl. Instrum. Meth.* **135**, 337–352 (1976).

70. Kohley, Z. et al. Modeling interactions of intermediate-energy neutrons in a plastic scintillator array with GEANT4. *Nucl. Instrum. Meth. Phys. Res. A* **682**, 59–65 (2012).

71. Lepage, G. P. & Brodsky, S. J. Exclusive processes in perturbative quantum chromodynamics. *Phys. Rev. D* **22**, 2157 (1980).

72. Chernyak, V. L. & Zhitnitsky, A. R. Asymptotic behavior of exclusive processes in QCD. *Phys. Rep.* **112**, 173–318 (1984).

73. Hansen, P. C. & O'Leary, D. The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM J. Sci. Comput.* **14**, 1487–1503 (1993).

Acknowledgements

All authors are members of the MINERvA Collaboration. This document was prepared by members of the MINERvA Collaboration using the resources of the Fermi National Accelerator Laboratory (Fermilab), a US Department of Energy, Office of Science, HEP User Facility. Fermilab is managed by Fermi Research Alliance, LLC (FRA), acting under contract no. DE-AC02-07CH11359. These resources included support for the MINERvA construction project, and support for construction also was granted by the United States National Science Foundation under award no. PHY-0619727 and by the University of Rochester. Support for participating scientists was provided by NSF and DOE (USA); by NSERC (Canada); by CAPES and CNPq (Brazil); by CoNaCyT (Mexico); by Proyecto Basal FB 0821, CONICYT PIA ACT1413, and Fondecyt 3170845 and 11130133 (Chile); by CONCYTEC (Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica), DGI-PUCP (Dirección de Gestión de la Investigación-

Pontificia Universidad Católica del Peru) and VRI-UNI (Vice-Rectorate for Research of National University of Engineering) (Peru); NCN Opus grant no. 2016/21/B/ST2/01092 (Poland); by Science and Technology Facilities Council (UK); by EU Horizon 2020 Marie Skłodowska-Curie Action. We thank the MINOS Collaboration for use of its near detector data. Finally, we thank the staff of Fermilab for support of the beamline, the detector and computing infrastructure.

Author information

Author notes

1. M. L. Moore

   Present address: Department of Physics, Stanford University, Stanford, CA, USA

2. M. V. Ascencio

   Present address: Iowa State University, Ames, IA, USA

3. A. Bashyal

   Present address: High Energy Physics/Center for Computational Excellence Department, Argonne National Laboratory, Lemont, IL, USA

4. G. Caceres

   Present address: Department of Physics and Astronomy, University of California at Davis, Davis, CA, USA

5. M. F. Carneiro

   Present address: Brookhaven National Laboratory, Upton, NY, USA

6. R. Fine

Present address: Los Alamos National Laboratory, Los Alamos, NM, USA

7. J. Kleykamp

   Present address: Department of Physics and Astronomy, University of Mississippi, Oxford, MS, USA

8. B. Messerly

   Present address: University of Minnesota, Minneapolis, MN, USA

9. B. Yaeggy

   Present address: Department of Physics, University of Cincinnati, Cincinnati, OH, USA

Authors and Affiliations

1. Department of Physics and Astronomy, University of Rochester, Rochester, NY, USA

   T. Cai, M. L. Moore, A. Olivier, A. Bercellie, A. Bodek, H. Budd, G. A. Díaz, R. Fine, S. Henry, J. Kleykamp, S. Manly, K. S. McFarland, G. N. Perdue, D. Ruterbories, M. Sultana & C. Wret

2. Department of Physics and Astronomy, York University, Toronto, Ontario, Canada

   T. Cai & D. A. Harris

3. AMU Campus, Aligarh, India

   S. Akhter, Z. Ahmad Dar & V. Ansari

4. Department of Physics, William & Mary, Williamsburg, VA, USA

   Z. Ahmad Dar, A. Filkins, M. Kordosky, O. Moreno, J. K. Nelson, E. Valencia & L. Zazueta

5. Sección Física, Departamento de Ciencias, Pontificia Universidad Católica del Perú, Lima, Peru

M. V. Ascencio & A. M. Gago

6. Department of Physics, Oregon State University, Corvallis, OR, USA

A. Bashyal, M. F. Carneiro, S. M. Gilligan, H. Schellman & N. H. Vaughan

7. Fermi National Accelerator Laboratory, Batavia, IL, USA

M. Betancourt, D. A. Harris, D. Jena, J. G. Morfín & G. N. Perdue

8. Campus León y Campus Guanajuato, Universidad de Guanajuato, Guanajuato, Mexico

J. L. Bonilla, J. Felix, E. Granados, O. Moreno, M. A. Ramírez & E. Valencia

9. University of Geneva, Geneva, Switzerland

A. Bravar

10. Centro Brasileiro de Pesquisas Físicas, Rio de Janeiro, Brazil

G. Caceres, M. F. Carneiro, H. da Motta & A. Lozano

11. Department of Physics, University of Notre Dame, Notre Dame, IN, USA

L. Fields

12. Physics Department, Tufts University, Medford, MA, USA

H. Gallagher, T. Le, W. A. Mann & V. S. Syrotenko

13. Department of Physics, University of Minnesota – Duluth, Duluth, MN, USA

R. Gran

14. Department of Physical Sciences, IISER Mohali, Mohali, India

    S. Jena

15. The Blackett Laboratory, Imperial College London, London, UK

    A. Klustová, A. V. Waldron & M. O. Wascko

16. Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA, USA

    D. Last, C. Mauger & M. A. Ramírez

17. Department of Physics, University of Warwick, Coventry, UK

    X.-G. Lu

18. Department of Physics, University of Oxford, Oxford, UK

    X.-G. Lu & K.-J. Plows

19. Massachusetts College of Liberal Arts, North Adams, MA, USA

    E. Maher

20. Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA, USA

    B. Messerly, D. Naples, V. Paolone & H. Su

21. Departamento de Física, Universidad Técnica Federico Santa María, Valparaíso, Chile

    J. Miller & B. Yaeggy

22. Department of Physics, University of Florida, Gainesville, FL, USA

    C. Nguyen & H. Ray

23. Rutgers, The State University of New Jersey, Piscataway, NJ, USA

R. D. Ransome

24. Facultad de Ciencias, Universidad Nacional de Ingeniería, Lima, Peru

C. J. Solano Salinas

25. School of Physical and Chemical Sciences, Queen Mary University of London, London, United Kingdom

A. V. Waldron

Contributions

The operation, Monte Carlo simulation and data analysis of the MINERvA Experiment are carried out by the MINERvA Collaboration with contributions from all collaborators listed as authors on this manuscript. The scientific results presented here have been presented to and discussed by the full collaboration, and all authors have approved the final version of the manuscript. T.C. and A.B. first proposed this analysis to the authors. D.R. developed the analysis tools with input from T.C. and A.O. T.C. performed the analysis of data and simulation described in this work. T.C. and M.L.M. performed the fits to the data. T.C. and K.S.M. wrote the manuscript.

Corresponding authors

Correspondence to T. Cai or K. S. McFarland.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Illustration of MINERvA detector.

A 3D illustration of the MINERvA detector (left), a flat cross-sectional portrayal (reproduced from ref .[11], right). The detector comprises 120 hexagonal modules, each consists of an Inner Detector (ID), side Electromagnetic Calorimeter (ECAL), and side Hadronic Calorimeter (HCAL) arranged outwards from the center. Horizontally along the beam direction in the ID region are four sub-detectors: nuclear target detector, active scintillator detector, the ECAL and the HCAL. Each scintillator plane in the detector is arranged in one of "X", "U", or "V" (bottom) orientations. A liquid helium tank is located upstream of the nuclear target detector. Downstream of the MINERvA detector is the magnetized MINOS[36] near detector acting as MINERvA's muon spectrometer.

Extended Data Fig. 2 Definition of $\delta\theta_P$ and $\delta\theta_R$, and the event distributions.

(top) Angular variable definitions. The variables $\delta\theta_R$ and $\delta\theta_P$ are defined according to the rotated reference frame ($(\hat{x},{\hat{y}}^{\prime},{\hat{z}}^{\prime})$). Distributions of (bottom left) $\delta\theta_P$ and (bottom right) $\delta\theta_R$ in the QELike sample normalized to bin width, after sideband fits. The vertical error bars around the data points represent 1 standard deviation due to statistical uncertainty.

Extended Data Fig. 3 Simulated event rate in $\delta\theta_P - \delta\theta_R$ plane in selected Q^2 analysis bins.

Heat map showing Monte Carlo event rate for CCE, CCQE, QELike 2p2h, and QELike Resonant interaction models in the $\delta\theta_R$-$\delta\theta_P$ plane at a few slices of Q^2. CCE events are concentrated around origin, while the CCQE events have broader spread. Both QELike 2p2h and resonant events show diffused structure going out to larger $\delta\theta_R$ and $\delta\theta_P$ regions. The color scales in the heat maps are different because of different event rates for each subsample.

Extended Data Fig. 4 Post-fit event rates and ratios in the validation regions.

The "Non-QE validation" region is shown separated into two sub-regions. "Non-QE Val. 1" spans $|\delta\theta_P| < 20°$, "Non-QE Val. 2" occupies $20° < |\delta\theta_P| < 55°$. The vertical error bars around the data points and the error band around the model prediction account for 1 standard deviation due to statistical uncertainty. The CCE signal and the regions used in the background fits are shown in Figs. 2 and 3, respectively.

Extended Data Fig. 5 Background constraint method tested on a neutrino sample.

(Top left) Signal and (top right) QE region distributions of the neutrino sample that looks for proton final states after fit. The vertical error bars around the data points and the error band around the model prediction account for 1 standard deviation due to statistical and systematic uncertainties. (Bottom) Application of $\pm$ 100% shift in 2p2h (gray band) on $\delta\theta_R$ for events in a $-10° < \delta\theta_P < 10°$ and $0.2\ (\mathrm{GeV}/c)^2 < Q^2 < 0.4\ (\mathrm{GeV}/c)^2$ slice. The analog to the CCE selection selects between $-10°$ and $10°$ in $\delta\theta_R$.

Extended Data Fig. 6 Signal event selection efficiency.

CCE signal efficiency as a function of Q^2. The inner error bar on each data point accounts for the statistical effect of a Poisson standard deviation, while

the full error bar account for all sources of systematic uncertainties.

Extended Data Fig. 7 Fractional uncertainties.

Fractional statistical and systematic uncertainties, as a function of Q^2 uncertainties. Systematic uncertainties in the “other” category, including the neutron and proton interaction uncertainties, are shown on the right. The neutron interaction systematic accounts for the neutron secondary interaction uncertainties in detector. The leading interaction channels, such as (nC, Bnp),(nC, 3α), and $((nC,{n}^{{\prime} }C\gamma)$), are assigned 10% to 15% uncertainties below a kinetic energy of 100 MeV.

Extended Data Fig. 8 Cross section and the Q^2-dependent flux cut.

(left) Measured cross-section and theory predictions, (right) Regions of neutrino energy and flux in signal selection at each Q^2. The inner error bars on the data points account for 1 standard deviation due to statistical uncertainty only, and the full error bars include contribution from all sources of systematic uncertainties.

Extended Data Table 1 Recoil energy cut

Supplementary information

Supplementary Information

Supplementary Tables 1–7.

Peer Review File

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation,

distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Cai, T., Moore, M.L., Olivier, A. *et al.* Measurement of the axial vector form factor from antineutrino–proton scattering. *Nature* **614**, 48–53 (2023). https://doi.org/10.1038/s41586-022-05478-3

- Received: 19 April 2022
- Accepted: 25 October 2022
- Published: 01 February 2023
- Issue Date: 02 February 2023
- DOI: https://doi.org/10.1038/s41586-022-05478-3

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05478-3

| Section menu | Main menu |

- Article
- Open Access
- Published: 01 February 2023

Field-linked resonances of polar molecules

- Xing-Yan Chen,
- Andreas Schindewolf,
- Sebastian Eppelt,
- Roman Bause,
- Marcel Duda,
- Shrestha Biswas,
- Tijs Karman,
- Timon Hilker,
- Immanuel Bloch &
- …
- Xin-Yu Luo

Show authors

Nature **volume 614**, pages 59–63 (2023)

- 1710 Accesses

- 26 Altmetric

- Metrics details

Abstract

Scattering resonances are an essential tool for controlling the interactions of ultracold atoms and molecules. However, conventional Feshbach scattering

resonances[1], which have been extensively studied in various platforms[1,2,3,4,5,6,7], are not expected to exist in most ultracold polar molecules because of the fast loss that occurs when two molecules approach at a close distance[8,9,10]. Here we demonstrate a new type of scattering resonance that is universal for a wide range of polar molecules. The so-called field-linked resonances[11,12,13,14] occur in the scattering of microwave-dressed molecules because of stable macroscopic tetramer states in the intermolecular potential. We identify two resonances between ultracold ground-state sodium–potassium molecules and use the microwave frequencies and polarizations to tune the inelastic collision rate by three orders of magnitude, from the unitary limit to well below the universal regime. The field-linked resonance provides a tuning knob to independently control the elastic contact interaction and the dipole–dipole interaction, which we observe as a modification in the thermalization rate. Our result provides a general strategy for resonant scattering between ultracold polar molecules, which paves the way for realizing dipolar superfluids[15] and molecular supersolids[16], as well as assembling ultracold polyatomic molecules.

Main

Ultracold polar molecules with tunable dipole moments provide a powerful platform for quantum simulations[17,18], quantum computation[19,20] and ultracold chemistry[21]. Scattering resonances are long-sought-after tools in these systems, which have been essential in ultracold-atom experiments to control the contact interaction and for creating strongly correlated quantum phases[22], as well as for producing ultracold diatomic molecules[1]. Independent control over contact and long-range interactions in ultracold molecules has been predicted to enable the realization of new quantum phenomena such as exotic self-bound droplets and supersolid quantum phases[16]. Moreover, measurements of scattering resonances provide an accurate benchmark for calculations of the molecular potential energy surface[3,21] and open a new route in controlled quantum chemistry[5].

A scattering resonance occurs when the scattering state strongly couples to a quasibound state. Based on whether the quasibound state is hosted by the

same or a different channel than the scattering channel, the resonance is categorized as a shape resonance or a Feshbach resonance, respectively. Shape and Feshbach resonances have been observed in atom–molecule and molecule–molecule collisions by scanning the collision energy by using molecular beams at kelvin and subkelvin temperatures[21,23,24,25,26]. In the ultracold (submicrokelvin) regime, scattering resonances are often induced by an external electromagnetic field that shifts the relative energy between the quasibound state and the scattering state[1]. Magnetically tunable Feshbach resonances have been observed in collisions between weakly bound Feshbach molecules[2,4] and recently between NaLi molecules in the spin-triplet ground state[6]. However, the magnetic tuning scheme essential to Feshbach resonances requires a non-zero electronic spin, and is, therefore, unlikely to find application for bialkali molecules in the spin-singlet ground state. The spin-singlet absolute ground state of bialkaline molecules is of special interest, as it is the only long-lived state in which the molecules feature strong electric dipole–dipole interactions (DDI). Moreover, Feshbach resonances are not expected to occur between ground-state molecules in the presence of nearly universal loss, owing to the high density of tetramer states near the collisional threshold and the loss mechanisms associated with collisional complexes[8,9,10]. A general method to realize collisional resonances of ultracold dipolar molecules, therefore, remains open.

Here we demonstrate a general approach to create such resonances in collisions between dipolar molecules by coupling them to so-called field-linked (FL) states[11,12]. These weakly bound states are induced by engineering an attractive well in the long-range intermolecular potential through microwave dressing[14,27]. Unlike in conventional resonances, in which an external field merely tunes an existing short-range quasibound state into resonance, the long-range FL states exist only in the presence of the microwave field. The sensitivity of the FL states to the microwave field leads to an unprecedented level of control over the intermolecular interaction. Here we demonstrate this tunability by observing two resonance branches in the inelastic scattering rate, the peak positions of which continuously shift with the microwave frequency and polarization. We further characterize the change of the thermalization rate caused by the diverging scattering length in a resonant collision channel.

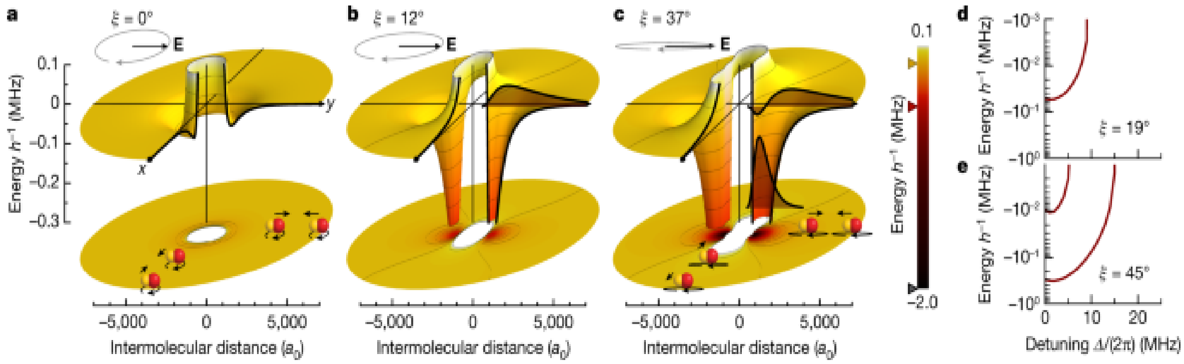
Interaction potential

Polar molecules possess a permanent dipole moment d_0 in their body-fixed frame. To induce a dipole moment in the laboratory frame, external fields need to be applied to mix different rotational states and to break the rotational symmetry. Here we use microwave dressing between the two lowest rotational states of the molecules to polarize them. The induced dipole moment follows the alternating current electric field as ${\bf{d}}({\rm{t}})=\bar{d}[{{\bf{e}}}_{+}(t)\cos \xi +{{\bf{e}}}_{-}(t)\sin \xi]$, where ξ describes the ellipticity of the microwave radiation and $\mathbf{e}_{\pm}(t)$ are the $\sigma^{\pm}$ polarization basis vectors. The time-averaged dipole moment $\bar{d}={d}_{0}/\sqrt{6(1+{(\varDelta /\varOmega)}^{2})}$ is tunable via the microwave detuning $\varDelta$ and the Rabi frequency, $\varOmega$. Because the rotational frequency of NaK at 5.643 GHz is much faster than all other dynamical timescales in the system, we consider the time-averaged DDI at long range[15]

$${U}_{{\rm{d}}{\rm{d}}}({\bf{r}})=\frac{{\bar{d}}^{2}}{8\pi {\epsilon }_{0}{r}^{3}}(3{\cos }^{2}\theta -1+3\sin 2\xi {\sin }^{2}\theta \cos 2\phi)$$
(1)

where $\mathbf{r} = (r, \theta, \varphi)$ is the relative position between the molecules in polar coordinates defined by the microwave wavevector. The microwave propagates in the z direction, and the microwave polarization ellipse has its major and minor axes in the y and x direction. Remarkably, the symmetry of the interaction can be manipulated by the microwave ellipticity, as illustrated in Fig. 1a–c. For $\xi = 0°$ (circular polarization) and $\xi = 45°$ (linear polarization), where U_{dd} resembles the typical DDI up to a constant prefactor[28]. In between linear and circular polarization, the interaction breaks the rotational symmetry along all directions.

Fig. 1: Interaction potentials and bound states of microwave-dressed ground-state molecules.



a–**c**, The cut-open three-dimensional surfaces illustrate the interaction potentials $U(\mathbf{r})$ including the *p*-wave centrifugal potential between two molecules in the *x*–*y* plane for different ellipticity angles, ξ, of the field polarization: $\xi = 0°$ (**a**), $\xi = 12°$ (**b**) and $\xi = 37°$ (**c**). Below, a projection of the same potential is shown. The interaction potential along the *z* direction is always repulsive (not shown). The microwave is on resonance ($\Delta = 0$). The shaded areas in **b** and **c** show the radial wavefunction of the bound states. The insets visualize the rotating electric field vector **E**, and sketch the interaction between the rotating dipoles colliding along the *x* or *y* direction. The markers on the colour bar denote the potential depths for the three cases. **d**,**e**, Coupled-channel calculations of the energy of the bound states as a function of Δ for values of $\xi = 19°$ (**d**) and $\xi = 45°$ (**e**). In all panels the Rabi frequency is set to $\Omega = 2\pi \times 10$ MHz.

Source data

As the molecules approach each other, microwave dressing induces an anisotropic van der Waals interaction $U_{\mathrm{vdW}} \propto 1/r^6$ (Methods)[15]. With blue-detuned microwave dressing, U_{vdW} is repulsive in all directions, which protects the molecules from loss processes at short range and reduces the inelastic cross-sections[14,29,30,31]. Such a shielding potential arises owing to an avoided crossing between the attractive and the repulsive branch of the DDI. In a semi-classical picture, this avoided crossing can be understood as the reorientation of colliding dipoles through DDI[31,32]. A similar flipping of the dipoles occurs between polar molecules in a direct current electric field[33,34], between Rydberg atoms[35] and between ions and Rydberg atoms[36].

The full interaction potential between two dressed molecules is the sum of the DDI potential and the van der Waals potential $U(\mathbf{r}) = U_{\rm dd}(\mathbf{r}) + U_{\rm vdW}(\mathbf{r})$. We can shape the interaction potential and control the scattering process, as illustrated in Fig. 1. Along the y axis, $U(\mathbf{r})$ resembles a Mie potential[15,37,38] with a characteristic length of about $10^3 a_0$. A deviation from circular polarization breaks the azimuthal symmetry of the DDI and enhances the depth of the potential well along the y axis, which becomes deep enough to support one or more bound states[13]. These bound states are the FL states, the properties of which strongly depend on the external fields. By tuning the binding energy of the FL state across the collisional threshold, for example, with microwave detuning as shown in Fig. 1d,e, FL resonances occur, which drastically alter the scattering properties between the molecules.

The low-energy scattering in $U(\mathbf{r})$ can be described by the associated partial-wave phase shifts, which are given by[39]

$${\delta }_{l{m}_{l}}({k})\approx {({k}{a}_{l{m}_{l}})}^{2l+1}+{k}{c}_{l{m}_{l}}{a}_{{\rm{d}}{\rm{d}}} \quad (2)$$

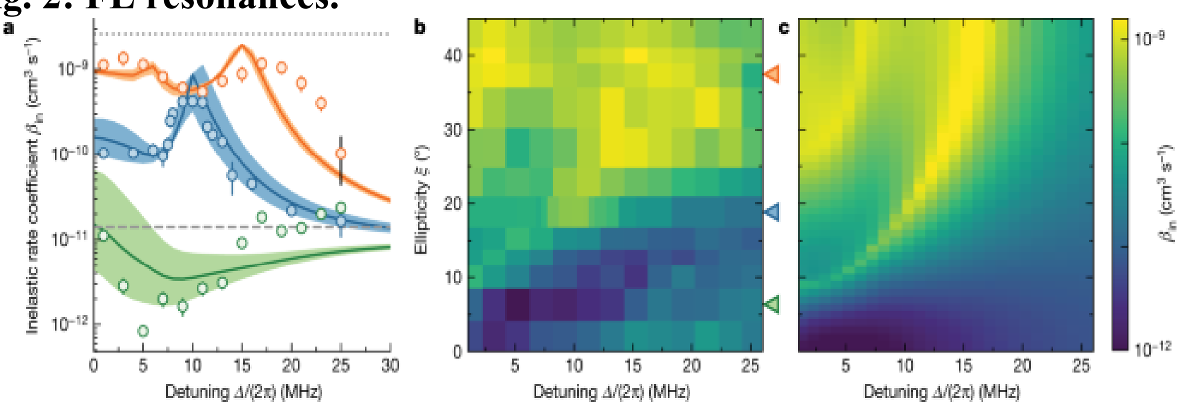
Here l and m_l are the angular momentum and its projection along the quantization axis, $\mathbf{k}$ is the relative wavevector, ${a}_{l{m}_{l}}$ and $a_{\rm dd}$ are the characteristic lengths associated with the contact interaction and the DDI respectively, and ${c}_{l{m}_{l}}$ denote partial-wave dependent prefactors for the dipolar scattering phase shifts. Note that the contact interaction is suppressed by the centrifugal barrier for $l \neq 0$ as $\mathbf{k} \to 0$, whereas the phase shift from the long-range DDI scales linearly with $\mathbf{k}$ in all partial waves, and is proportional to the dipolar length ${a}_{{\rm{d}}{\rm{d}}}=\mu {\bar{d}}^{2}/4\pi {\epsilon }_{0}{\hslash }^{2}$, where $\mu = m/2$ is the reduced mass of the molecule. The FL resonances provide a tuning knob for ${a}_{l{m}_{l}}$ in the resonant channel. By reducing $U_{\rm vdW}$, FL resonances can occur at any desired dipole moment up to ${d}_{0}/\sqrt{6}$, thus realizing independent control of the contact interaction and the DDI.

Resonance map

We map out the resonances by measuring the inelastic rate coefficient β_{in} of collisions between the dressed molecules. The optically trapped ground-state ^{23}Na^{40}K molecules with nuclear spin projections $(m_{i,\text{Na}}, m_{i,\text{K}}) = (3/2, -4)$ are formed from an ultracold atomic mixture by means of magnetoassociation and subsequent stimulated Raman adiabatic passage (STIRAP) at a magnetic offset field of 72.35 G (ref. [40]). For most measurements, the temperature T of the molecular ensemble is 230 nK and the initial average density n_0 is about $5 \times 10^{11}\ \text{cm}^{-3}$. Next, the microwave is ramped on in 100 μs to dress the molecules. After a variable hold time, the remaining molecules are released from the optical trap and we determine the number of molecules and the temperature from the time-of-flight images.

As we tune the ellipticity of the microwave from circular to linear, up to two FL states emerge from the dressed potential. Figure 2a shows exemplary loss rate coefficients for three different ellipticities. At $\xi = 6(2)°$, the potential is too shallow to support a bound state, thus no resonance is observed. In this regime, the inelastic collision is suppressed by the shielding potential at small detunings[31]. For $\xi = 19(2)°$, the interaction potential supports a single bound state near zero microwave detuning, leading to enhanced inelastic scattering at $\Delta \approx 2\pi \times 10$ MHz. For $\xi = 37(2)°$, the potential becomes deep enough to support two bound states, leading to two resonance peaks.

Fig. 2: FL resonances.



a, Inelastic collision rate coefficient β_{in} between microwave-dressed NaK molecules as a function of the microwave detuning Δ for various microwave polarizations with the ellipticity angle $\xi = 6(2)°$ (green), 19(2)° (blue) and 37(2)° (orange) at the Rabi frequency $\Omega \approx 2\pi \times 10$ MHz. The solid lines

show the corresponding theory calculations. The shaded regions show the calculations within the uncertainty of ξ. The grey dashed line denotes the theoretical universal value of β_{in} and the grey dotted line denotes the single-channel unitarity limit. The coloured error bars show the standard deviation of the fit results and the black error bar illustrates the exemplary common systematic uncertainty. **b**,**c**, Colour density map of the experiment data (**b**) and the theory calculation (**c**) of the inelastic rate coefficient as a function of microwave detuning and ellipticity. The triangles on the right axis of **b** mark ellipticity for the data shown in **a**.

Source data

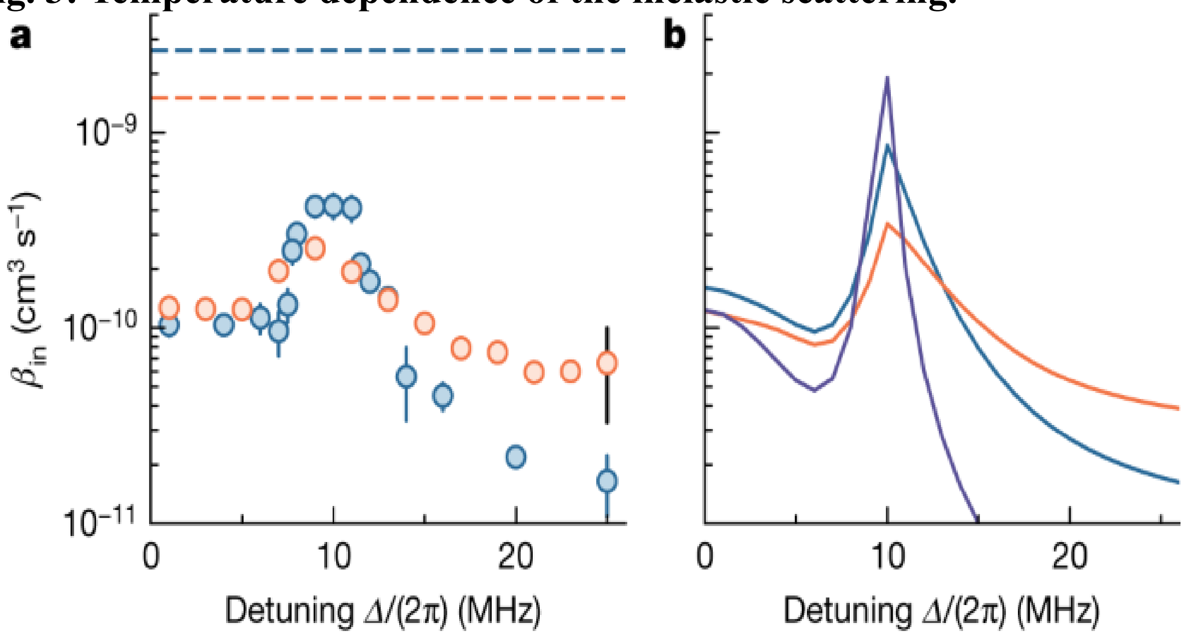
One special feature of the FL resonances is their sensitivity to external fields. We show that the resonance position continuously changes with the microwave parameters by mapping out the two resonance branches while varying the microwave detuning and polarization. Figure 2b shows two branches of FL resonances, starting at $\xi \approx 10°$ and $\xi \approx 32°$. As the polarization ellipticity ξ increases, less DDI is needed to support the bound states and the resonances, therefore, shift to larger detuning. However, the global inelastic rate coefficient increases as the polarization becomes more elliptical, because of the increased coupling to the other dressed states[28]. Overall our measurements show good agreement with our theory predictions (Fig. 2c). We attribute the broadening and shift of the resonance peaks compared to the theory to an increase of the Rabi frequency as we scan the detuning (Methods). These systematic errors affect mostly the FL resonances near linear polarization, where the potential depth is more sensitive to the relative detuning.

Temperature dependence of the inelastic scattering

The temperature dependence of the inelastic scattering rate varies with the detuning. At large detuning where the DDI is reduced, the inelastic scattering rate is universal[41] and scales for identical fermions linearly with temperature. At small detuning, the scattering enters the semi-classical regime for which β_{in} is independent of the temperature[42]. On the scattering resonance, the collision rate has a temperature dependence that is

reminiscent of the unitarity limit[5,32], whereas the loss remains substantially smaller than this limit because of shielding. Meanwhile, the width of the resonance feature is broadened by thermal averaging. As a consequence, for temperatures as high as 700 nK, the resonance becomes less visible as shown in Fig. 3. When the collision energy becomes lower than the centrifugal barrier of the interaction potential, the resonance peak would be further narrowed owing to the increased lifetime of the quasibound state. Therefore, reaching ultracold temperatures is crucial for the observation of FL resonances.

Fig. 3: Temperature dependence of the inelastic scattering.



a,**b**, Experimental results (**a**) and calculations (**b**) of the inelastic collision rate coefficient β_{in} as a function of the microwave detuning Δ at 700 nK (orange), 230 nK (blue) and 20 nK (purple). The coloured error bars show fitting errors, and the black error bar additionally contains the common systematic uncertainty. The solid lines are coupled-channel calculations and the dashed lines are the unitarity limit at the corresponding temperatures. The molecules are dressed by microwaves with ellipticity $\xi = 19(2)°$ and Rabi frequency $\Omega \approx 2\pi \times 10$ MHz.

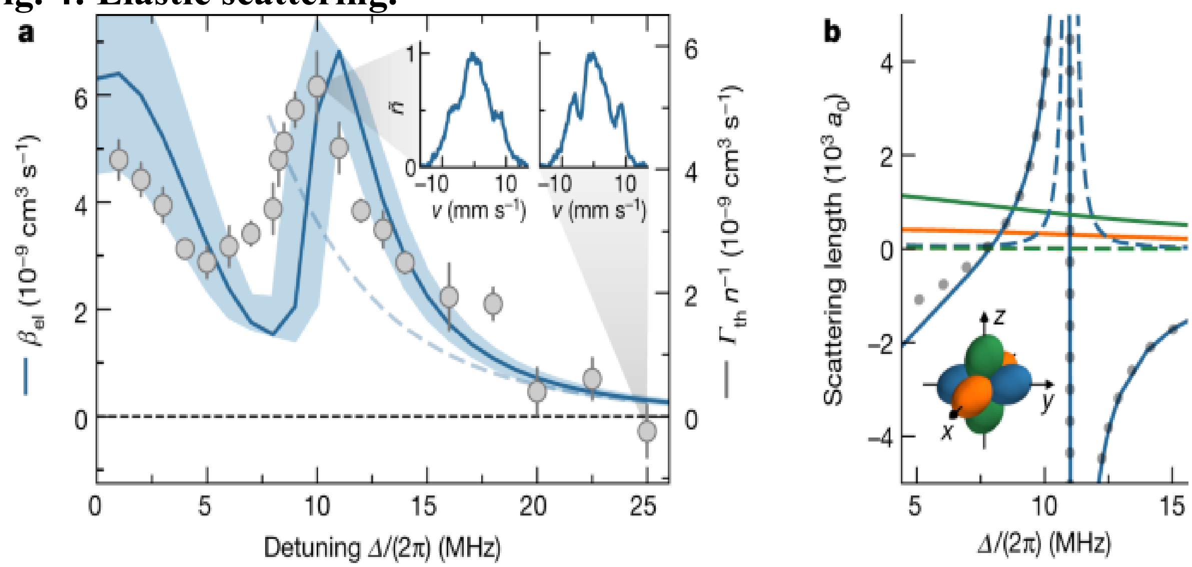
Source data

Elastic scattering

Scattering resonances are associated not only with enhanced losses of the molecules, but more importantly, the ability to control elastic scattering. With FL resonances, we can tune the elastic scattering rate while keeping the inelastic rate small.

We characterize the effect of the FL resonances on the elastic collision rate by measuring the thermalization rate of the samples. This is commonly done by quenching the trapping confinement in one dimension and observing the global cross-dimensional thermalization[43]. However, for small Δ/Ω our samples are in the hydrodynamic regime, in which the global thermalization rate is limited by the trapping frequencies[31]. Instead, we perturb the momentum distribution of the molecular cloud by pulsing on an optical-lattice beam for 300 µs. The lattice pulse diffracts some molecules and sends them to collide along the *y* axis, defined as the long-axis of the microwave field (Fig. 1), along which the DDI is most attractive. Fast local thermalization smears out the diffraction pattern that is formed in momentum space during the lattice pulse (insets in Fig. 4a). From the contrast of the diffraction pattern we can estimate the thermalization rate Γ_{th} (Methods).

Fig. 4: Elastic scattering.



a, The grey data points show the thermalization rate Γ_{th} normalized by the mean in situ density n as a function of the microwave detuning Δ at an ellipticity angle $\xi = 19(2)°$ and a Rabi frequency $\Omega \approx 2\pi \times 10$ MHz. The

temperature is 230 nK. The error bars are the standard error of the mean of 7 to 16 repetitions. For comparison, the solid blue line shows the corresponding theory calculation of the elastic collision rate coefficient β_{el}. The uncertainty of ξ is taken into account by the shaded area. The dashed line is the Born approximation of the background collision rate coefficient, which holds for detunings $\Delta \gtrsim 2\pi \times 8$ MHz (Methods). The insets show the normalized linear density \(\widetilde{n}\) along the lattice axis averaged over ten repetitions as a function of the molecule velocity v in the lattice direction for $\Delta = 2\pi \times 10$ MHz and $2\pi \times 25$ MHz. **b**, The coupled-channel calculations of the energy-dependent scattering length with the same microwave parameters and a fixed collision energy of $k_B T$, with $T = 230$ nK. The solid (dashed) lines are the real (imaginary) part of the scattering lengths in the channel p_x (orange), p_y (blue) and p_z (green). The dotted line is a fit of equation (3). The inset illustrates the partial waves of the scattering channels. k_B is the Boltzmann constant.

Source data

We observe that the measured thermalization rate follows a similar trend as the calculated value of the elastic collision rate coefficient, β_{el} (Fig. 4a). Besides the contribution from the DDI, which scales with \({\bar{d}}^{2}\) and decreases as Δ increases, a clear resonance feature is visible around the FL resonance. The shift of the resonance between the experiment data and theory is within the systematic uncertainty of the ellipticity. The average number of elastic collisions that is needed per particle to reach thermalization is so far unknown in the regime of elliptical microwave polarization. From the comparison between the measured Γ_{th}/n and β_{el} from our coupled-channel calculations, we find that this factor is close to 1 under the present experimental conditions.

The observed elastic scattering rate has contributions from both contact interaction and long-range DDI. As shown in equation (2), the DDI contributes in multiple channels, whereas the contact interaction only has a notable contribution in the resonant channel, owing to its unfavourable scaling with the wavevector. Although such an interplay between contact interaction and DDI limits the change in the total elastic scattering rate, the underlying scattering length, however, shows divergent behaviour in the

resonant channel. Figure 4b shows the energy-dependent scattering length ${\mathop{a}\limits^{ \sim }}_{{{\rm{l}}{\rm{m}}}_{l}}({k})=-\,\tan {\delta }_{{{\rm{l}}{\rm{m}}}_{l}}({k})/{k}$ (refs. [41,44]) for the three p-wave channels at the average collision energy. The real (imaginary) part of the scattering length corresponds to the elastic (inelastic) scattering. The FL resonance occurs in the p_y channel where the interaction is most attractive. The corresponding scattering length shows a resonance feature, in which the real part can be tuned to large positive or negative values, whereas the imaginary part remains small. The ratio of elastic-to-inelastic collisions is about ten on the resonance, and can be further enhanced at smaller ellipticity, in which the resonance shifts towards higher Rabi frequencies (Methods).

A simple analytic formula that describes the resonant elastic scattering length is given by

$${\mathop{a}\limits^{ \sim }}_{1y}={a}_{{\rm{d}}{\rm{d}}}\left(-\frac{1+3\sin 2\xi }{10}\right)\left(1+\frac{{\varDelta }^{\ast }({k})}{\varDelta -{\varDelta }_{0}({k})}\right) \quad (3)$$

where $\varDelta_0(k)$ and $\varDelta^*(k)$ denote the position and the width of the resonance. The width $\varDelta^*(k) \propto k^2$ follows the scaling of the p-wave contact interaction. For the collision energy considered here, we extract $\varDelta_0 \approx 2\pi \times 10.99$ MHz and $\varDelta^* \approx 2\pi \times 3.29$ MHz from the fit to the coupled-channel calculations. The resonance position $\varDelta_0(\mathbf{k})$ also has a weak energy dependence. As a consequence, the resonance position with thermal averaging is broadened and slightly shifted towards lower detuning. These thermal effects, however, will be greatly suppressed in a degenerate Fermi gas, in which the scattering predominantly occurs near the sharp Fermi energy[15].

Discussion

Field-linked resonances provide a new universal tool to control the collisions between ultracold polar molecules. These resonances occur as long as the Rabi frequency is sufficiently large, such that the interaction potential is

deep enough to support the bound states. Quantitative predictions of FL resonances only require knowledge of the mass, the dipole moment and the rotational structure of the individual molecules as well as their loss rate at short range. This is in stark contrast to molecular collisions involving close contact between the molecules, in which a large number of collision channels are involved and the existing knowledge of the potential energy surface are too imprecise to predict the number of bound states, let alone their position. Precise knowledge on the FL states also makes them useful as intermediate states in photoassociation spectroscopy to probe the short-range potential[11].

The control over the scattering length opens up new possibilities to investigate many-body physics with both contact interaction and DDI. In a degenerate Fermi gas, the resonant interaction facilitates realization of dipolar superfluidity[27,45,46]. Specifically, pairing between molecules is enhanced because of the presence of the FL bound state. Therefore, the critical temperature for Bardeen–Cooper–Schrieffer superfluidity increases drastically near the FL resonance, that is, to about 14% of the Fermi temperature for NaK molecules[15]. The anisotropic nature of such a dipolar superfluid gives rise to new quantum phenomena such as gapless superfluidity[47] and topological $p_x + ip_y$ symmetry[48]. In a Bose–Einstein condensate (BEC), independent control over the *s*-wave scattering length and dipolar length has led to the observation of self-bound droplets and the formation of supersolids in magnetic atoms[49]. Making use of FL resonance with bosonic polar molecules, the dipolar lengths of which are orders of magnitude larger than magnetic atoms, would enable the study of such exotic phenomena in entirely unexplored regimes[16].

The observed resonances also demonstrate the existence of the FL states, a new set of exotic long-range polyatomic molecular states. Tetramer molecules with approximately twice the dipole moment of the individual diatomic molecules could potentially be created by adiabatically ramping the microwave field across a FL resonance or by radio-frequency association. Those composite bosonic tetramers are expected to be long lived at small binding energies[12] and could be collisionally stable owing to the shielding between the constituent molecules. Below the critical temperature, a

tetramer gas could form a BEC[50] and may lead to a new crossover from a dipolar Bardeen–Cooper–Schrieffer superfluid to a BEC of tetramers.

Conclusion

We have observed a new type of universal scattering resonance between ultracold microwave-dressed polar molecules that is associated with FL tetramer bound states in the long-range potential well. The resonances are highly tunable via microwave power, frequency and polarization, which makes them a versatile tool for controlling molecular interactions. Because the FL states are insensitive to species-dependent short-range interactions, the FL resonance is applicable to a wide range of polar molecules. Our results provide a general route to strongly interacting molecular gases and open up new possibilities to investigate new quantum many-body phenomena and to produce long-lived dipolar tetramer molecules.

Methods

Microwave field generation

We use a dual-feed square-waveguide antenna to generate a microwave field with tunable polarization. The waveguide is fabricated from copper-coated glass-fibre-reinforced epoxy laminate. The inner dimensions of the waveguide are 33 × 33 × 58 mm. The width of the waveguide is chosen such that the cut-off frequency of the transverse electric (TE_{10}) mode is below the rotational transition frequency of the molecules at 5.64 GHz and that the microwave field strength at the position of the molecules is optimal. To achieve impedance matching, the two feeds are 13 mm long, which is a quarter wavelength in free space. They are placed orthogonal to each other, and 22 mm away from the back plate of the waveguide, which is about a quarter wavelength in the waveguide. Each feed produces a close to linearly polarized electric field at the position of the molecules, which is about 25 mm away from the top of the waveguide. When the field strengths of these subfields are balanced, tuning the relative phase of the feeds allows straightforward tuning of the field polarization. A relative phase of approximately 90° (−90°) results in a σ^+ (σ^-) polarized field, whereas

approximately 0° (180°) produces a linearly polarized field along the y (x) direction.

The microwave setup is sketched in Extended Data Fig. 1. A Rohde & Schwarz SMA100B signal generator (using a yttrium–iron–garnet oscillator) with the noise suppression option SMAB-B711 is used as microwave source. To independently control the two feeds of the antenna, the microwave is divided by a power splitter into two paths. Each path includes a voltage-controlled attenuator to balance the subfields, a 10-W amplifier (KUHNE electronic KU PA 510590-10 A) and a mechanical phase shifter (SHX BPS-S-6-120) for differential phase control.

Calibration of the field polarization

To characterize the polarization of the microwave field, we initially measure its π, σ^+ and σ^- component in the frame of the magnetic offset field. We do this at low microwave power by measuring the Rabi coupling to excited rotational states with different projections m_J of the rotational quantum number J on the magnetic field axis [31]. A direct measurement of the σ^+ and σ^- component at large microwave power is impracticable, as the Rabi frequency is then orders of magnitude larger than the Zeeman splitting of the m_J states. When we change the relative phase ϕ between the feeds, each field component, and thus each Rabi frequency, individually oscillates with a period of 360° owing to the interference between the imperfect subfields, as shown in Extended Data Fig. 2a. We fit these oscillations with the function

$$\varOmega (\varphi)=\sqrt{{\varOmega }_{1}^{2}+{\varOmega }_{2}^{2}+2{\varOmega }_{1}{\varOmega }_{2}\cos (\varphi +{\varphi }_{0})} \quad (4)$$

with the fit parameters Ω_1, Ω_2 and ϕ_0, which define the contributions of the individual feed. The offset phases ϕ_0 have an uncertainty of 2.9°, which we attribute to the hysteresis and imperfect tuning of the mechanical phase shifters. Note, for the calibration measurements presented in Extended Data Fig. 2a, the power balance between the feeds was tuned to minimize the σ^-

component around $\phi = 90°$. The finite ellipticity of the subfields causes the field strengths at other angles of ϕ to be unbalanced, so that we do not get pure σ^+ polarization at $\phi = -90°$ and the phase values that provide linear polarization shift away from 0° and 180°.

We calculate the ellipticity ξ from the fitted Rabi frequencies for each individual relative phase, as shown in Extended Data Fig. 2b. At high microwave power, the electric field of the microwave replaces the magnetic offset field as the quantization axis, so that ξ is defined in the frame of the microwave field. Because the relative phase between the measured field components is unknown, we can only deduce the limits for the tilt angle of the microwave wavevector with respect to the magnetic offset field, which causes a systematic uncertainty of ξ. Close to σ^+ polarization the tilt angle is only around 10°, so that the uncertainty of ξ is dominated by the above-mentioned uncertainties of ϕ_0.

When switching from the calibration measurements at low microwave power to the measurements at high power, we have to consider the non-linearity of the amplifiers close to their saturation power. This shifts the power balance between the two antenna feeds when we increase the microwave power and thereby changes ξ. We reestablish the power balance by tuning one voltage-controlled attenuator while optimizing the shielding at $\phi = 90°$, where we initially minimized ξ. We find that we need to compensate the relative power by 10%. In addition, we also observe a small variation of the relative phase and power as we scan the microwave detuning. The variation in the relative phase leads to a systematic uncertainty of the ellipticity on the same order as the contribution from the uncertainty of ϕ_0. The variation in Rabi frequency could broaden the resonance feature, especially for ellipticities close to linear polarization, as observed in Fig. 2.

Inelastic collision rate coefficient

The inelastic collision rate coefficients β_{in} are experimentally determined from the time evolution of the measured molecule number N and the temperature T by numerically solving the differential equations[51]

$$\frac{{\rm{d}}N}{{\rm{d}}t}=\left(-{\beta }_{{\rm{in}}}n-{\varGamma }_{1}\right)N$$
(5)
$$\frac{{\rm{d}}T}{{\rm{d}}t}=h$$
(6)

with the average density

$$n=\frac{N}{8\sqrt{{{\rm{\pi }}}^{3}{k}_{{\rm{B}}}^{3}{T}^{3}/{m}^{3}{\bar{\omega }}^{6}}}$$
(7)

where $\bar{\omega }$ is the geometric mean trapping frequency and $\varGamma_1$ is the one-body loss rate. We assume both the heating rate and the two-body loss rate coefficient to be constant during the loss process. As the overall heating is no more than 50%, the fitted values of $\beta_{\rm in}$ do not significantly change when we instead assume a linear temperature dependence of the rate coefficient.

The comparison between loss measurements on the scattering resonance and away from resonance are shown in Extended Data Fig. 3a. To limit the number of free fit parameters, we determine $\varGamma_1 = 0.53(2)$ Hz in independent measurements at low densities, as shown in Extended Data Fig. 3b.

Fast thermalization measurements

The momentum distribution of the molecular cloud is disturbed by pulsing on a one-dimensional optical lattice for $t_{\rm lat} = 300$ μs. Subsequently the microwave power is ramped down in 100 μs and the trapping confinement is turned off. After 10 ms time of flight, the momentum distribution is imaged. The lattice is approximately aligned along the *y* direction. Its lattice constant is $a_{\rm lat} = 532$ nm and the lattice depth is $88E_{\rm r}$, where ${E}_{{\rm{r}}}={h}^{2}/(8m{a}_{{\rm{lat}}}^{2})$ is the lattice recoil energy. The pulse duration was chosen to optimize the contrast of the resulting interference pattern. Note that the pulse is short compared to the trap frequencies 2π × (82, 58, 188) Hz of the background confinement, so that cross-talk between

the momentum distribution and the real-space density is small. Also two-body loss, even on resonance, is negligible on this time scale.

To estimate the degree of thermalization $c_{\rm th}$ we make the simplifying assumption that the momentum distribution after the lattice pulse can be described by

$$\widetilde{n}(v)={c}_{{\rm{th}}}{\widetilde{n}}_{{\rm{th}}}(v)+(1-{c}_{{\rm{th}}}){\widetilde{n}}_{0}(v)$$
(8)

where ${\widetilde{n}}_{{\rm{th}}}(v)$ is a Gaussian (that is, thermal) distribution and ${\widetilde{n}}_{0}(v)$ is the undisturbed interference pattern (Extended Data Fig. 4a). We determine ${\widetilde{n}}_{0}(v)$ by averaging over 16 images of the interference pattern in the absence of microwave-induced interactions. We further assume that the interference pattern decays exponentially, such that the thermalization rate is given by

$${\varGamma }_{{\rm{th}}}=-\,{\rm{ln}}(1-{c}_{{\rm{th}}})/{t}_{{\rm{lat}}}$$
(9)

Given the average number of elastic collisions that is required to reach thermal equilibrium $N_{\rm col}$ and the average relative velocity $v_{\rm rel}$ of the colliding molecules, the elastic scattering cross-section is given by

$${\sigma }_{{\rm{el}}}=\frac{{N}_{{\rm{col}}}{\varGamma }_{{\rm{th}}}}{n{v}_{{\rm{rel}}}}$$
(10)

Assuming that the thermalization is mainly driven by elastic collisions between molecules in the side peaks and molecules in the main peak of the diffraction pattern, the average relative velocity can be approximated as

$${v}_{{\rm{rel}}}=\sqrt{{\bar{v}}^{2}+{h}^{2}/{(m{a}_{{\rm{lat}}})}^{2}}$$
(11)

where $\bar{v}=\sqrt{16{k}_{{\rm{B}}}T/({\rm{\pi }}m)}$ is the thermally averaged collision velocity in the undisturbed sample. The expression of N_{col} is so far unknown in the regime of elliptical microwave polarization. We can, however, set an upper limit by assuming $\xi = 0$. In that case, $N_{col} \approx 4$ for a tilt angle of the microwave field of 10° (refs. [34,43]). A comparison of this simplified model with our predictions of ${\beta }_{{\rm{el}}}=\bar{v}{\sigma }_{{\rm{el}}}$ is shown in Extended Data Fig. 4b.

Born approximation

The elastic scattering rate (Fig. 4a) from the DDI is determined by the long-range U_{dd}. In the low-energy regime $E\lesssim {\hbar }^{2}/\mu {a}_{{\rm{dd}}}^{2}$, the elastic scattering rate can be obtained via the Born approximation ${\beta }_{{\rm{el}}}={\sigma }_{{\rm{el}},{\rm{Born}}}\bar{v}$, where

$${\sigma }_{{\rm{el}},{\rm{Born}}}=\frac{16{\rm{\pi }}}{15}(1+3{\sin }^{2}2\xi){a}_{{\rm{dd}}}^{2}$$
(12)

is the elastic cross-section[42].

We can also obtain the dipolar scattering lengths from the Born approximation. Because U_{dd} is symmetric under reflection along the three Cartesian axes, the p_x, p_y and p_z channels are decoupled and the main contributions are the elastic scattering within each channel. The corresponding scattering lengths are given by

$${\widetilde{a}}_{1x,{\rm{Born}}}=-\,\frac{1}{10}(1-3\sin 2\xi){a}_{{\rm{dd}}}$$
(13)
$${\widetilde{a}}_{1y,{\rm{Born}}}=-\,\frac{1}{10}(1+3\sin 2\xi){a}_{{\rm{dd}}}$$
(14)
$${\widetilde{a}}_{1z,{\rm{Born}}}=\frac{1}{5}{a}_{{\rm{dd}}}$$
(15)

The scattering length given in equation (3) is ${\widetilde{a}}_{1y,{\rm{Born}}}$ plus the contribution from the FL resonance.

Resonance near circular microwave polarization

The elastic-to-inelastic collision ratio near the FL resonances can be improved with better shielding near the circular microwave polarization and with lower temperature. The ratio is then about 900 at the maximum elastic scattering rate and about 130 at the scattering resonance, as shown in Extended Data Fig. 5. Under these conditions, the FL resonance occurs at a much higher Rabi frequency compared to the observed resonances at more elliptical polarizations. However, this is still realistic to achieve by using an improved antenna design with increased microwave power.

Induced van der Waals interaction

Reference[15] provides an analytical formula for the induced van der Waals interaction

$$\begin{array}{l}{U}_{{\rm{vdW}}}({\bf{r}})=\frac{35{C}_{{\rm{6i}}}}{4{r}^{6}}{\sin }^{2}\theta \left({\cos }^{2}\theta +1\right.\\ \,-\left.\,2\sin 2\xi {\cos }^{2}\theta \cos 2\phi -{\sin }^{2}2\xi {\sin }^{2}\theta {\cos }^{2}2\phi \right)\end{array}$$
(16)

where the induced C_{6i} coefficient is given by

$${C}_{{\rm{6i}}}=\frac{{d}_{0}^{4}}{1120{{\rm{\pi }}}^{2}{\epsilon }_{0}^{2}\varOmega {\left(1+{(\varDelta /\varOmega)}^{2}\right)}^{3/2}}$$
(17)

For $\Omega = 2\pi \times 10$ MHz and zero detuning, $C_{6i} = 1.1 \times 10^{-72}$ kg m^8 s^{-1} is orders of magnitude larger than the direct isotropic van der Waals coefficient $C_6 = -4.9 \times 10^{-74}$ kg m^8 s^{-1} (ref. [52]). The induced C_{6i} reduces as the detuning increases and can be comparable to C_6 for $\Delta/\Omega > 2.7$.

Coupled-channel calculations

We perform coupled-channel scattering calculations by using the framework described in detail in refs. [28,29]. Here we summarize the numerical details of these calculations.

The NaK molecules are described as rigid rotors with the rotational states $J = 0, 1$. Interactions of the molecules with elliptically polarized microwaves are included as described in ref. [53]. The wavefunction for the relative motion is expanded in partial waves $L = 1, 3, 5$. Hyperfine interactions are not included as these were previously found to have a negligible effect[31] when operating at sufficiently large magnetic fields[29,53]. The colliding molecules interact with one another through the DDI, and undergo short-range loss, which is modelled as a capture boundary condition imposed at $r = 20a_0$. We propagate the coupled-channel equations outwards to $r = 10^6 a_0$ and match the solution to the scattering boundary conditions. This yields the scattering matrix, from which collision cross-sections and rate coefficients are determined and thermally averaged by using an energy grid of 21 energies that are logarithmically spaced between $0.03k_BT$ and $32k_BT$.

Binding energies of the FL bound states are calculated as follows. First, we compute adiabatic potentials by diagonalizing the Hamiltonian described above excluding the radial kinetic energy for fixed values of the molecule–molecule separation r. On each adiabatic potential curve, we compute bound states by using a sinc-function discrete variable representation[54]. We find that the position of the zero-energy bound states computed in this approximation agree well with the resonance positions, indicating that one can think of the FL bound states as living on a single adiabatic potential curve. Both resonances found here are supported by the lowest adiabatic potential, that is, the second resonance corresponds to a radial vibrational excitation, rather than an angular excitation. Interaction potentials that are shown in Fig. 1 are similarly computed as adiabatic potential curves, except that these are computed for a fixed orientation of the intermolecular axis relative to the microwave polarization, rather than by using a partial wave expansion.

Note, the coupled-channel calculations have no free parameters.

Data availability

The experimental data that support the findings of this study are available from the corresponding author upon reasonable request. Source data are provided with this paper.

Code availability

All relevant codes are available from the corresponding author upon reasonable request.

References

1. Chin, C., Grimm, R., Julienne, P. & Tiesinga, E. Feshbach resonances in ultracold gases. *Rev. Mod. Phys.* **82**, 1225–1286 (2010).

2. Chin, C. et al. Observation of Feshbach-like resonances in collisions between ultracold molecules. *Phys. Rev. Lett.* **94**, 123201 (2005).

3. Yang, H. et al. Observation of magnetically tunable Feshbach resonances in ultracold $^{23}Na^{40}K + ^{40}K$ collisions. *Science* **363**, 261–264 (2019).

4. Wang, F., Ye, X., Guo, M., Blume, D. & Wang, D. Observation of resonant scattering between ultracold heteronuclear Feshbach molecules. *Phys. Rev. A* **100**, 042706 (2019).

5. Son, H. et al. Control of reactive collisions by quantum interference. *Science* **375**, 1006–1010 (2022).

6. Park, J. J., Lu, Y.-K., Jamison, A. O., Tscherbul, T. & Ketterle, W. A Feshbach resonance in collisions between ultracold ground state molecules. Preprint at http://arxiv.org/abs/2208.05557 (2022).

7. Weckesser, P. et al. Observation of Feshbach resonances between a single ion and ultracold atoms. *Nature* **600**, 429–433 (2021).

8. Mayle, M., Ruzic, B. P. & Bohn, J. L. Statistical aspects of ultracold resonant scattering. *Phys. Rev. A* **85**, 062712 (2012).

9. Christianen, A., Karman, T. & Groenenboom, G. C. Quasiclassical method for calculating the density of states of ultracold collision complexes. *Phys. Rev. A* **100**, 032708 (2019).

10. Liu, Y. & Ni, K.-K. Bimolecular chemistry in the ultracold regime. *Annu. Rev. Phys. Chem.* **73**, 73–96 (2022).

11. Avdeenkov, A. V. & Bohn, J. L. Linking ultracold polar molecules. *Phys. Rev. Lett.* **90**, 043006 (2003).

12. Avdeenkov, A. V., Bortolotti, D. C. E. & Bohn, J. L. Field-linked states of ultracold polar molecules. *Phys. Rev. A* **69**, 012710 (2004).

13. Ticknor, C. & Bohn, J. L. Long-range scattering resonances in strong-field-seeking states of polar molecules. *Phys. Rev. A* **72**, 032717 (2005).

14. Lassablière, L. & Quéméner, G. Controlling the scattering length of ultracold dipolar molecules. *Phys. Rev. Lett.* **121**, 163402 (2018).

15. Deng, F. et al. Effective potential and superfluidity of microwave-dressed polar molecules. Preprint at https://arxiv.org/abs/2210.13253 (2022).

16. Schmidt, M., Lassablière, L., Quéméner, G. & Langen, T. Self-bound dipolar droplets and supersolids in molecular Bose-Einstein condensates. *Phys. Rev. Res.* **4**, 013235 (2022).

17. Carr, L. D., DeMille, D., Krems, R. V. & Ye, J. Cold and ultracold molecules: science, technology and applications. *New J. Phys.* **11**, 055049 (2009).

18. Baranov, M. A., Dalmonte, M., Pupillo, G. & Zoller, P. Condensed matter theory of dipolar quantum gases. *Chem. Rev.* **112**, 5012–5061 (2012).

19. DeMille, D. Quantum computation with trapped polar molecules. *Phys. Rev. Lett.* **88**, 067901 (2002).

20. Ni, K.-K., Rosenband, T. & Grimes, D. D. Dipolar exchange quantum logic gate with polar molecules. *Chem. Sci.* **9**, 6830–6838 (2018).

21. Balakrishnan, N. Perspective: ultracold molecules and the dawn of cold controlled chemistry. *J. Chem. Phys.* **145**, 150901 (2016).

22. Bloch, I., Dalibard, J. & Zwerger, W. Many-body physics with ultracold gases. *Rev. Mod. Phys.* **80**, 885–964 (2008).

23. Skodje, R. T. et al. Resonance-mediated chemical reaction: F + HD → HF + D. *Phys. Rev. Lett.* **85**, 1206–1209 (2000).

24. Henson, A. B., Gersten, S., Shagam, Y., Narevicius, J. & Narevicius, E. Observation of resonances in Penning ionization reactions at sub-kelvin temperatures in merged beams. *Science* **338**, 234–238 (2012).

25. Chefdeville, S. et al. Observation of partial wave resonances in low-energy O_2–H_2 inelastic collisions. *Science* **341**, 1094–1096 (2013).

26. de Jongh, T. et al. Imaging the onset of the resonance regime in low-energy NO–He collisions. *Science* **368**, 626–630 (2020).

27. Cooper, N. R. & Shlyapnikov, G. V. Stable topological superfluid phase of ultracold polar fermionic molecules. *Phys. Rev. Lett.* **103**, 155302 (2009).

28. Karman, T., Yan, Z. Z. & Zwierlein, M. Resonant and first-order dipolar interactions between ultracold $^1\Sigma$ molecules in static and microwave electric fields. *Phys. Rev. A* **105**, 013321 (2022).

29. Karman, T. & Hutson, J. M. Microwave shielding of ultracold polar molecules. *Phys. Rev. Lett.* **121**, 163401 (2018).

30. Anderegg, L. et al. Observation of microwave shielding of ultracold molecules. *Science* **373**, 779–782 (2021).

31. Schindewolf, A. et al. Evaporation of microwave-shielded polar molecules to quantum degeneracy. *Nature* **607**, 677–681 (2022).

32. Yan, Z. Z. et al. Resonant dipolar collisions of ultracold molecules induced by microwave dressing. *Phys. Rev. Lett.* **125**, 063401 (2020).

33. Matsuda, K. et al. Resonant collisional shielding of reactive molecules using electric fields. *Science* **370**, 1324–1327 (2020).

34. Li, J.-R. et al. Tuning of dipolar interactions and evaporative cooling in a three-dimensional molecular quantum gas. *Nat. Phys.* **17**, 1144–1148 (2021).

35. Hollerith, S. et al. Quantum gas microscopy of Rydberg macrodimers. *Science* **364**, 664–667 (2019).

36. Zuber, N. et al. Observation of a molecular bond between ions and Rydberg atoms. *Nature* **605**, 453–456 (2022).

37. Stone, A. J. *The Theory of Intermolecular Forces (International Series of Monographs on Chemistry)*, Vol. 32 (Clarendon Press, 1997).

38. Du, R., Qi, R. & Zhang, P. Solutions of the Schrödinger equation for anisotropic dipole-dipole interaction plus isotropic van der Waals interaction. Preprint at https://arxiv.org/abs/2207.09377 (2022).

39. Sadeghpour, H. R. et al. Collisions near threshold in atomic and molecular physics. *J. Phys. B: At. Mol. Opt. Phys.* **33**, R93–R140 (2000).

40. Duda, M. et al. Transition from a polaronic condensate to a degenerate Fermi gas of heteronuclear molecules. Preprint at https://arxiv.org/abs/2111.04301 (2021).

41. Idziaszek, Z. & Julienne, P. S. Universal rate constants for reactive collisions of ultracold molecules. *Phys. Rev. Lett.* **104**, 113202 (2010).

42. Bohn, J. L., Cavagnero, M. & Ticknor, C. Quasi-universal dipolar scattering in cold and ultracold gases. *New J. Phys.* **11**, 055039 (2009).

43. Wang, R. R. W. & Bohn, J. L. Anisotropic thermalization of dilute dipolar gases. *Phys. Rev. A* **103**, 063320 (2021).

44. Hutson, J. M. Feshbach resonances in ultracold atomic and molecular collisions: threshold behaviour and suppression of poles in scattering lengths. *New J. Phys.* **9**, 152 (2007).

45. You, L. & Marinescu, M. Prospects for *p*-wave paired Bardeen-Cooper-Schrieffer states of fermionic atoms. *Phys. Rev. A* **60**, 2324–2329 (1999).

46. Baranov, M. A., Mar'enko, M. S., Rychkov, V. S. & Shlyapnikov, G. V. Superfluid pairing in a polarized dipolar Fermi gas. *Phys. Rev. A* **66**, 013606 (2002).

47. Gorshkov, A. V. et al. Tunable superfluidity and quantum magnetism with ultracold polar molecules. *Phys. Rev. Lett.* **107**, 115301 (2011).

48. Levinsen, J., Cooper, N. R. & Shlyapnikov, G. V. Topological $p_x + ip_y$ superfluid phase of fermionic polar molecules. *Phys. Rev. A* **84**, 013603 (2011).

49. Chomaz, L. et al. Dipolar physics: a review of experiments with magnetic quantum gases. Preprint at https://arxiv.org/abs/2201.02672 (2022).

50. Greiner, M., Regal, C. A. & Jin, D. S. Emergence of a molecular Bose–Einstein condensate from a fermi gas. *Nature* **426**, 537–540 (2003).

51. Ni, K.-K. et al. Dipolar collisions of polar molecules in the quantum regime. *Nature* **464**, 1324–1328 (2010).

52. Żuchowski, P. S., Kosicki, M., Kodrycka, M. & Soldán, P. van der Waals coefficients for systems with ultracold polar alkali-metal molecules. *Phys. Rev. A* **87**, 022706 (2013).

53. Karman, T. & Hutson, J. M. Microwave shielding of ultracold polar molecules with imperfectly circular polarization. *Phys. Rev. A* **100**, 052704 (2019).

54. Colbert, D. T. & Miller, W. H. A novel discrete variable representation for quantum mechanical reactive scattering via the *S*-matrix Kohn method. *J. Chem. Phys.* **96**, 1982–1991 (1992).

Acknowledgements

We thank T. Shi for stimulating discussions and providing the analytical formula for the interaction potential, J. Hutson, R. R. W. Wang, Y. Bao and H. Adel for stimulating discussions, and C. Buchberg and M. Hani for the cooperation on the development of the waveguide antenna. We gratefully acknowledge support from the Max Planck Society, the European Union (PASQuanS grant no. 817482) and the Deutsche Forschungsgemein–schaft under Germany's Excellence Strategy–EXC-2111–390814868 and under grant no. FOR 2247 and MCQST seed funding program. A.S. and T.H. acknowledge funding from the Max Planck Harvard Research Center for Quantum Optics.

Funding

Open access funding provided by Max Planck Society.

Author information

Author notes

1. These authors contributed equally: Xing-Yan Chen, Andreas Schindewolf

Authors and Affiliations

1. Max-Planck-Institut für Quantenoptik, Garching, Germany

   Xing-Yan Chen, Andreas Schindewolf, Sebastian Eppelt, Roman Bause, Marcel Duda, Shrestha Biswas, Timon Hilker, Immanuel Bloch & Xin-Yu Luo

2. Munich Center for Quantum Science and Technology, München, Germany

   Xing-Yan Chen, Andreas Schindewolf, Sebastian Eppelt, Roman Bause, Marcel Duda, Shrestha Biswas, Timon Hilker, Immanuel Bloch & Xin-Yu Luo

3. Institute for Molecules and Materials, Radboud University, Nijmegen, The Netherlands

   Tijs Karman

4. Fakultät für Physik, Ludwig-Maximilians-Universität, München, Germany

   Immanuel Bloch

Contributions

All authors contributed substantially to the work presented in this manuscript. X.-Y.C., A.S. and S.E. carried out the experiments and together with R.B., M.D. and S.B. improved the experimental setup. X.-Y.C., A.S. and S.E. analysed the data. T.K. performed the theoretical calculations. T.H., I.B. and X.-Y.L. supervised the study. All authors worked on the interpretation of the data and contributed to the final manuscript.

Corresponding author

Correspondence to Xin-Yu Luo.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Microwave setup.

a, Electronic setup used to generate and control the microwave field. Individually, the two antenna feeds produce mainly linear polarized fields parallel to each feed, respectively. The voltage-controlled attenuators are used to balance the fields and to adiabatically ramp the field intensity. The phase shifters allow to tune the polarization. **b**, Half-section view of the waveguide antenna that shows the inside of the waveguide and the transition from the coaxial cable (red jacket) to the feed.

Extended Data Fig. 2 Calibration of the field polarization.

a, The Rabi frequencies of rotational σ^+ (green), π (blue), and σ^- (orange) transitions at low microwave power as a function of the phase shift ϕ between the antenna feeds. The error bars show the fitting error of the Rabi oscillations. The solid lines are fits to equation (4). **b**, Ellipticity of the microwave field in the frame of the microwave. The data points show the ellipticity angle ξ calulated from the data in **a**. The error bars denote the uncertainty of ξ that originates from the projection from the frame of the magnetic offset field to the frame of the microwave field due to the unknown phase relation between the three field components. The gray band is calculated from the fit functions shown in **a** and its width considers the uncertainty of the microwave orientation and the uncertainty of the offset phases ϕ_0.

[Source data](#)

[Extended Data Fig. 3 One- and two-body loss.](#)

a, Example molecule loss at $\xi \approx 19°$ with detuning on resonance $\Delta = 2\pi \times 10$ MHz (bright) and away from resonance $\Delta = 2\pi \times 25$ MHz (dark). The lines are fits to the differential-equation model. **b**, Loss at low initial densities from which the one-body loss rate Γ_1 is determined. The line is an exponential fit function. The error bars show the standard deviation for repeated experiments.

[Source data](#)

[Extended Data Fig. 4 Thermalization model.](#)

a, The blue line shows the diffraction pattern $\tilde{n}(v)$ from a single experiment run at $\Delta = 2\pi \times 10$ MHz. The red line is a fit to equation ([8](#)). The gray curve describes the thermalized part $c_{th}\tilde{n}_{th}(v)$. **b**, The same measurement as in Fig. [4a](#) but here a simplified model is used to determine β_{el} from the experimental data. The error bars show the standard error of the mean of 7–16 repetitions.

[Source data](#)

[Extended Data Fig. 5 Field-linked resonance near circular microwave polarization.](#)

Coupled-channel calculations of the elastic (orange) and inelastic (blue) scattering rate coefficient at $T = 20$ nK, $\xi = 1°$, and $\Delta = 2\pi \times 1$ MHz.

[Source data](#)

Supplementary information

[Peer Review File](#)

Source data

Source Data Fig. 1

Source Data Fig. 2

Source Data Fig. 3

Source Data Fig. 4

Source Data Extended Data Fig. 2

Source Data Extended Data Fig. 3

Source Data Extended Data Fig. 4

Source Data Extended Data Fig. 5

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Chen, XY., Schindewolf, A., Eppelt, S. *et al.* Field-linked resonances of polar molecules. *Nature* **614**, 59–63 (2023). https://doi.org/10.1038/s41586-022-05651-8

- Received: 11 October 2022
- Accepted: 13 December 2022
- Published: 01 February 2023
- Issue Date: 02 February 2023
- DOI: https://doi.org/10.1038/s41586-022-05651-8

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

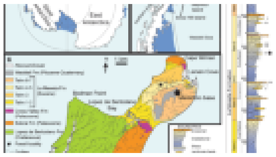
Provided by the Springer Nature SharedIt content-sharing initiative

Ultracold molecules find the sweet spot for collisions

- Sebastian Will
- Tanya Zelevinsky

Nature News & Views 01 Feb 2023

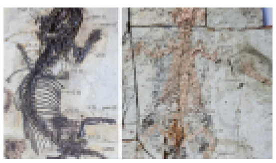
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05651-8



First fossil frog from Antarctica: implications for Eocene high latitude climate conditions and Gondwanan cosmopolitanism of Australobatrachia

23 April 2020

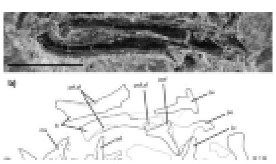
Thomas Mörs, Marcelo Reguero & Davit Vasilyan



Fossoriality and evolutionary development in two Cretaceous mammaliamorphs

07 April 2021

Fangyuan Mao, Chi Zhang, … Jin Meng



A tiny new Middle Triassic stem-lepidosauromorph from Germany: implications for the early evolution of lepidosauromorphs and the Vellberg fauna

20 February 2020

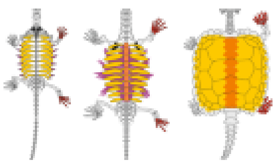
Gabriela Sobral, Tiago R. Simões & Rainer R. Schoch



[A fast-growing basal troodontid (Dinosauria: Theropoda) from the latest Cretaceous of Europe]

01 March 2021

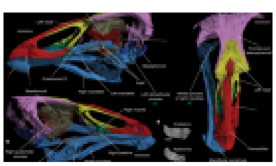
Albert G. Sellés, Bernat Vila, … Àngel Galobart



[Microanatomy of the stem-turtle Pappochelys rosinae indicates a predominantly fossorial mode of life and clarifies early steps in the evolution of the shell]

18 July 2019

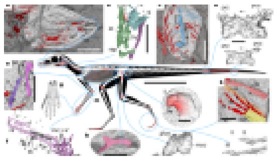
Rainer R. Schoch, Nicole Klein, … Hans-Dieter Sues



[Late Cretaceous neornithine from Europe illuminates the origins of crown birds]

18 March 2020

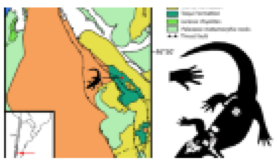
Daniel J. Field, Juan Benito, … Daniel T. Ksepka



[Scleromochlus and the early evolution of Pterosauromorpha]

05 October 2022

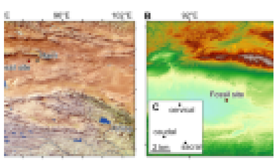
Davide Foffa, Emma M. Dunne, … Paul M. Barrett



[New transitional fossil from late Jurassic of Chile sheds light on the origin of modern crocodiles]

22 July 2021

Fernando E. Novas, Federico L. Agnolin, … Marcelo P. Isasi



[The first dinosaurs from the Early Cretaceous Hami Pterosaur Fauna, China]

12 August 2021

Xiaolin Wang, Kamila L. N. Bandeira, … Alexander W. A. Kellner

- Article
- Open Access
- Published: 25 January 2023

Triassic stem caecilian supports dissorophoid origin of living amphibians

- Ben T. Kligman ,
- Bryan M. Gee ,
- Adam D. Marsh ,
- Sterling J. Nesbitt ,
- Matthew E. Smith ,
- William G. Parker &
- …
- Michelle R. Stocker

Show authors

Nature **volume 614**, pages 102–107 (2023)

- 11k Accesses
- 539 Altmetric
- Metrics details

Abstract

Living amphibians (Lissamphibia) include frogs and salamanders (Batrachia) and the limbless worm-like caecilians (Gymnophiona). The estimated Palaeozoic era gymnophionan–batrachian molecular divergence[1] suggests a major gap in the record of crown lissamphibians prior to their earliest fossil occurrences in the Triassic period[2,3,4,5,6]. Recent studies find a monophyletic Batrachia within dissorophoid temnospondyls[7,8,9,10], but the absence of pre-Jurassic period caecilian fossils[11,12] has made their relationships to batrachians and affinities to Palaeozoic tetrapods controversial[1,8,13,14]. Here we report the geologically oldest stem caecilian—

a crown lissamphibian from the Late Triassic epoch of Arizona, USA—extending the caecilian record by around 35 million years. These fossils illuminate the tempo and mode of early caecilian morphological and functional evolution, demonstrating a delayed acquisition of musculoskeletal features associated with fossoriality in living caecilians, including the dual jaw closure mechanism[15,16], reduced orbits[17] and the tentacular organ[18]. The provenance of these fossils suggests a Pangaean equatorial origin for caecilians, implying that living caecilian biogeography reflects conserved aspects of caecilian function and physiology[19], in combination with vicariance patterns driven by plate tectonics[20]. These fossils reveal a combination of features that is unique to caecilians alongside features that are shared with batrachian and dissorophoid temnospondyls, providing new and compelling evidence supporting a single origin of living amphibians within dissorophoid temnospondyls.

Main

Of the nine tetrapod lineages surviving from the Triassic to the present day[21], caecilians have the most depauperate fossil record, with only 11 total occurrences[22]; of these, only *Rubricacaecilia monbaroni*[23] and *Eocaecilia micropodia*[11,12] represent unambiguous stem caecilians. The estimated Permo–Carboniferous origin of caecilians leaves a gap exceeding 70 million years between putative Palaeozoic relatives and *Eocaecilia*[1]. The absence of a pre-Jurassic caecilian record provides little evidence informing the pattern of morphological transformations leading to the specialized caecilian body plan, the timing and pattern of caecilian origins and diversification, the functional and ecological origins of extant caecilians, and caecilian palaeobiogeography. Furthermore, this gap has resulted in longstanding disagreement regarding the relationships of living amphibian groups to each other and to other tetrapods with multiple mutually exclusive hypotheses proposed[8,14]. With the discovery of *Gerobatrachus hottoni*[7], an early Permian dissorophoid bearing a combination of batrachian and amphibamiform features, the monophyly of Batrachia nested within amphibamiform dissorophoids reached near-consensus opinion[8], demonstrating the crucial nature of new fossil evidence to questions of lissamphibian origins. Despite the improved understanding of batrachian

origins, the origins of Lissamphibia remain contentious, now hinging on the relationships of caecilians to batrachians and Palaeozoic tetrapods[1,8,13,14]. Therefore, consensus on lissamphibian origins can be resolved only with the addition of new caecilian fossils filling the morphological gap between *Eocaecilia* and Palaeozoic tetrapods.

Here we approach such consensus by reporting the discovery of a new stem caecilian from a multitaxic microvertebrate and macrovertebrate bonebed in the Upper Triassic Chinle Formation of Petrified Forest National Park (PEFO), Arizona, USA (Extended Data Figs. 1 and 2). This material represents the most abundant caecilian-bearing fossil locality known, with at least 76 individuals consisting of isolated three-dimensional skeletal elements that we infer to belong to the same taxon, including elements from the upper and lower jaws, and postcrania (Supplementary Information, section 1).

Systematic palaeontology

Lissamphibia Haeckel, 1866

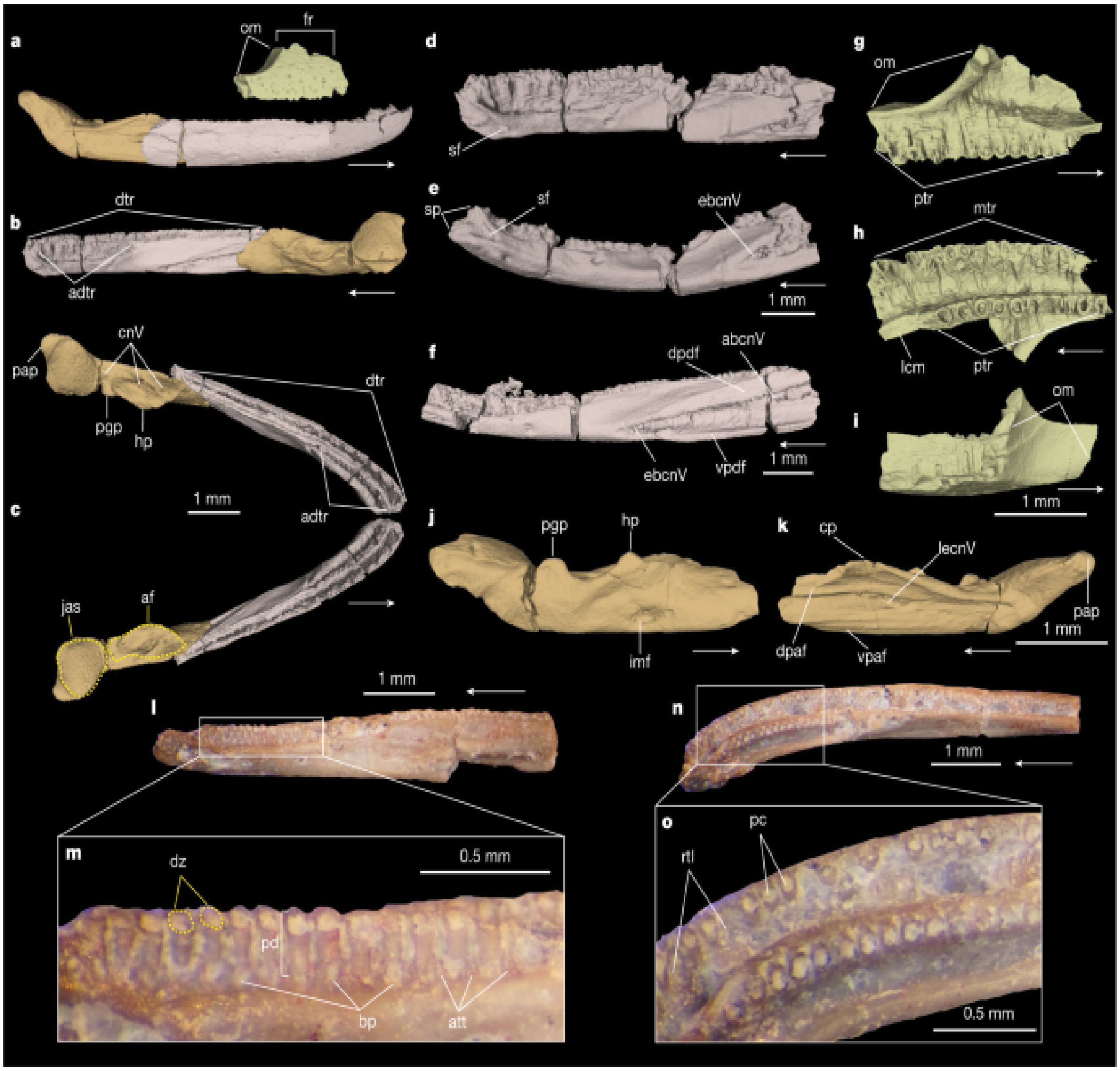
Gymnophionomorpha Marjanović and Laurin, 2008

Funcusvermis gilmorei gen. et sp. nov.

Etymology. *Funcus*, Latinized form of the English word funky (funk is an upbeat, rhythmic form of dance music); *vermis*, worm (Latin); in honour of the 1972 song Funky Worm from the album *Pleasure* by the Ohio Players. The species name honours N. Gilmore, collections manager at the Academy of Natural Sciences of Drexel University in Philadelphia, PA, USA.

Holotype. PEFO 43891, right pseudodentary (Fig. 1 and Extended Data Figs. 3 and 4), accessioned at Petrified Forest National Park, Arizona, USA.

Fig. 1: Digital renderings of holotype, paratype, and referred specimens of *F. gilmorei*.



a–**c**, Composite reconstruction of craniomandibular elements in lateral (**a**), medial (**b**) and dorsal (**c**) views. **d**,**e**, Holotype right pseudodentary (PEFO 43891) in medial and ventral views. **f**, Paratype right pseudodentary (PEFO 46284) in medial view. **g**–**i**, Referred left maxillopalatine (PEFO 46481) in medial (**g**), ventral (**h**) and dorsal (**i**) views. **j**,**k**, Referred left pseudoangular (PEFO 46480) in medial and lateral views. **l**–**o**, Paratype right pseudodentary (PEFO 45800) in medial (**l**; expanded view in **m**) and dorsal (**n**; expanded view in **o**) views. abcnV, alveolar branch cranial nerve V; adtr, adsymphyseal tooth row; af, adductor fossa; att, attachment tissue; bp, basal pore; cnV, cranial nerve V insertions; cp, coronoid process; dpaf, dorsal pseudoangular facet; dpdf, dorsal pseudodentary facet; dtr, dentary tooth row; dz, dividing zone; ebcnV, external branch cranial nerve V; fr, facial

ramus; hp, hamate process; imf, intramandibular foramen; jas, jaw articulation surface; lcm, lateral choanal margin; lecnV, lateral exit cranial nerve V; mtr, maxillary tooth row; om, orbital margin; pap, posterior pseudoangular process; pc, pulp cavity; pd, pedicel; pgp, preglenoid process; ptr, palatal tooth row; rtl, replacement tooth locus; sf, symphyseal foramen; sp, symphyseal prongs; vpaf, ventral pseudoangular facet; vpdf, ventral pseudodentary facet. Arrows indicate anterior direction.

Paratypes. PEFO 44432, PEFO 45800 and PEFO 46284 (all right pseudodentaries; Fig. 1 and Extended Data Figs. 3 and 4). Additional paratypes are listed in Supplementary Information, section 1.

Referred material. PEFO 46481, left maxillopalatine (Fig. 1 and Extended Data Fig. 4); PEFO 46480, left pseudoangular (Fig. 1 and Extended Data Fig. 3); PEFO 45810 (postatlantal vertebra), PEFO 43811 (right femur) (Extended Data Fig. 3). Additional referred specimens are listed in Supplementary Information, section 1.

Type locality and horizon. PFV 456, Thunderstorm Ridge, PEFO, Arizona, USA (Extended Data Fig. 2), within the upper Blue Mesa Member, Chinle Formation (Late Triassic: Norian); 223.036 ± 0.059 Ma (ref. [24] to 218.08 ± 0.037 Ma (ref. [25]), or ~221 Ma (ref. [26]); Adamanian estimated holochron[27]).

Diagnosis. A gymnophionomorph diagnosed by the following unique combination of features found in the holotype and paratype and referred specimens (asterisk denotes autapomorphies): symphyseal foramen* and notch subdividing the mandibular symphysis into medial and lateral processes*; at least 50 and at least 22 tooth pedicels in the dentary and adsymphyseal tooth rows, respectively. Further diagnosed by features found in referred specimens: co-ossified maxilla and palatine (compound maxillopalatine); palatal dentition of maxillopalatine terminated anteriorly by the lateral choanal margin*; maxillopalatine without osteological correlate of the tentacular organ*; absence of internal and retroarticular processes of the pseudoangular*; jaw articulation surface of pseudoangular formed by a subcircular flat pad; pseudoangular bearing a dorsally exposed adductor chamber occupying more than 30% of pseudoangular length*; three cranial nerve V insertions in pseudoangular*; femur present. Differential diagnosis in Supplementary Information, section 2.

Phylogenetic relationships

We tested the relationships of *Funcusvermis gilmorei* in a modified dataset[6] of 63 terminal taxa including stem tetrapods, stem and crown amniotes, and temnospondyl amphibians including stereospondyls and lissamphibians (Methods). Using both maximum parsimony and Bayesian inference optimality criteria (Methods), our phylogenetic analyses robustly support *Funcusvermis* as the earliest-diverging gymnophionomorph, sister taxon to the clade including *Eocaecilia, Rubricacaecilia* and Gymnophiona. All analyses unambiguously recovered a monophyletic Lissamphibia nested within amphibamiform dissorophoids, with *Gerobatrachus* and *Doleserpeton annectens* as successive outgroups to Lissamphibia (Fig. 3 and Extended Data Figs. 5–7). Our parsimony analysis recovered Lissamphibia consisting of a polytomous trichotomy of Gymnophionomorpha, Batrachia and Albanerpetontidae (Extended Data Fig. 5), whereas our Bayesian analysis recovered Lissamphibia consisting of a sister group relationship between Batrachia and a clade comprised of a sister group relationship between Gymnophionomorpha and Albanerpetontidae (Extended Data Fig. 7). The varying position of albanerpetontids in these and other recent analyses[28] highlights the ghost lineage from 150 million years ago (Ma) preceding their earliest occurrences in the Middle Jurassic epoch[29] as an outstanding gap obscuring conclusive resolution of relationships amongst major lissamphibian lineages. These results suggest that the caecilian-like anatomy in *Chinlestegophis jenkinsi* (a Late Triassic diminutive burrowing stereospondyl also found in the Chinle Formation[14]) is convergent with that of gymnophionomorphs such as *Eocaecilia* because of adaptations facilitating fossoriality (further discussed in Supplementary Information, section 3 and Extended Data Figs. 8–10).

Origins of the lissamphibian jaw apparatus

Funcusvermis indicates that many features of the lissamphibian mandibular ramus appeared initially in amphibamiform dissorophoids and were later lost or modified in batrachians, albanerpetontids and gymnophionomorphs (Fig. 3). All dentition in *Funcusvermis* (Fig. 1) is pedicellate—the oldest known example of this distinctive tooth form in crown Lissamphibia—reinforcing

hypotheses that pedicellate teeth are derived in amphibamiform dissorophoids[10], conserved in gymnophionomorphs and batrachians[30], and lost in albanerpetontids[29]. The rod-like pseudodentary of *Funcusvermis* resembles that of *Eocaecilia*[12] and the dentary of *Doleserpeton*[9] in the presence of tightly packed homodont tooth pedicels in parallel labial (dentary) and lingual (adsymphyseal) rows. The symphyseal foramen, of similar form and position to those of albanerpetontids[31], suggests that the Meckel's cartilage never ossified at the mandibular symphysis, probably a conservation of the ancestral condition of temnospondyls[32], and differing from the ossified condition of this element that forms a closed mandibular symphysis in batrachians and other gymnophionomorphs. As in *Doleserpeton*[9], a vertical notch bisects the mandibular symphysis between the anterior termini of the dentary and adsymphyseal tooth rows forming medial and lateral processes in *Funcusvermis* (Fig. 1 and Extended Data Fig. 3); these are similar to the more pronounced symphyseal prongs of albanerpetontids (for example, in refs. [28,29,31]), indicating that this feature may be ancestral to Lissamphibia and later lost in Batrachia and the common ancestor of *Eocaecilia* and Gymnophiona.

In *Funcusvermis*, the presence of 22 or more teeth in the adsymphyseal tooth row is similar to the more than 20 teeth reported in *Eocaecilia*[12], suggesting a transformation of the condition exhibited in *Doleserpeton* (5–7 teeth[9]) through distal expansion via addition of new teeth. In living caecilians, the lingual tooth row forms embryonically on a distinct anlage that later ossifies to the medial surface of the dentary forming the medial part of the mandibular symphysis and the lingual tooth row of adult caecilians[33]. Recent identifications of a dorsally facing tooth-bearing adsymphyseal (equivalent to the parasymphyseal (plate)) medial to (and separate from) the dentary at the mandibular symphysis in early branching tetrapods[34,35], juvenile temnospondyls[36] and dissorophoid temnospondyls[37] suggests that in taxa that appear to bear a lingual tooth row at the mandibular symphysis of the 'dentary' (for example, *Doleserpeton*, *Funcusvermis* and other gymnophionomorphs), the 'dentary' is actually composed of a tooth-bearing adsymphyseal (forming the lingual tooth row) co-ossified lingually to the dentary, and not a coronoid as previously thought[14,33] (Fig. 3 and Extended

Data Fig. 8; see Supplementary Information, section 2 for discussion of adsymphyseal homology).

The pseudoangular of *Funcusvermis* is highly similar to the postdentary morphology of dissorophoids exemplified by the amphibamid *Doleserpeton*[9] (Fig. 3 and Extended Data Fig. 3); as in *Doleserpeton*, batrachians, and albanerpetontids, *Funcusvermis* lacks retroarticular and internal processes, suggesting their initial acquisition in the common ancestor of *Eocaecilia* and Gymnophiona. The absence of the retroarticular process and presence of a dorsally facing adductor fossa (insertion site of the m. adductor mandibulae complex (mAM)) occupying more than 30% of pseudoangular length (Supplementary Table 1) in the pseudoangular of *Funcusvermis* (Fig. 1) illuminate a major transformation from the ancestral lissamphibian condition to the unique musculoskeletal architecture of living gymnophionans. Jaw closure driven primarily by the mAM is ancestral for tetrapods, and is retained in batrachians, albanerpetontids and *Funcusvermis*, differing from the condition of all other gymnophionomorphs, which exhibit the distinctive caecilian dual jaw closure mechanism[15] (DJCM). The DJCM is driven primarily by the hyobranchial muscle *m. interhyoidus posterior* (mIHP), and secondarily by the mAM; the mIHP inserts onto the ventral side of the retroarticular process and extends posteroventrally, acting as a first-order lever causing the anterior component of the lower jaw to pivot upwards with respect to the quadrate during jaw closure[15]. Acquisition of DJCM is hypothesized to be an adaptation for fossoriality: the mIHP contribution to bite force allows for reduction of the mAM and therefore compaction of the skull roof[15,16], a suite of transformations shown to be acquired by the common ancestor of *Eocaecilia* and Gymnophiona. Although the skull roof of *Funcusvermis* is unknown, absence of the DJCM and our phylogenetic results suggest that it probably retained the plesiomorphic condition of cheek emargination (gymnokrotaphy, as in *Gerobatrachus*, batrachians, albanerpetontids and presumably the common ancestor of Lissamphibia) to accommodate the mAM, rather than a closed skull roof with large interpterygoid vacuities[13] (stegokrotaphy, as in dissorophoids).

The obtuse angle of the orbital margin in the *Funcusvermis* maxillopalatine (Fig. 1) may suggest the presence of large orbits as in dissorophoids, batrachians and albanerpetontids (differing from the reduced orbits of other

gymnophionomorphs); however, the incomplete orbital margin in the single maxillopalatine specimen (PEFO 46481) prohibits conclusive assessment of this feature. The orbital margin of *Funcusvermis* lacks a tentacular fossa or aperture (osteological correlates for the chemosensory tentacle organ[18]), suggesting its absence in early gymnophionomorphs and later derivation by the common ancestor of *Eocaecilia* and Gymnophiona[12]. The presence of a co-ossified maxilla and palatine (maxillopalatine) in *Funcusvermis* is shared with gymnophionans and differs from that of amphibamiforms, albanerpetontids and batrachians, evidence of maxillopalatine consolidation early in gymnophionomorph evolution; however, these bones are possibly separate in *Eocaecilia*[12] and *Rubricacaecilia*[23]. Ventrally, the maxillopalatine of *Funcusvermis* bears parallel maxillary and palatal rows of tightly packed pedicellate teeth of similar size to those in the pseudodentary, seemingly intermediate between the condition of these dentitions in *Doleserpeton* and *Eocaecilia*, sharing an anterior truncation of the palatal tooth row by the internal nares with the former, and mesiodistal distal extension (through addition of new teeth) of the palatal row with the latter. A comprehensive comparative description of the *Funcusvermis* skull and postcranial elements is included in Supplementary Information, section 2.

Evolution of caecilian fossoriality

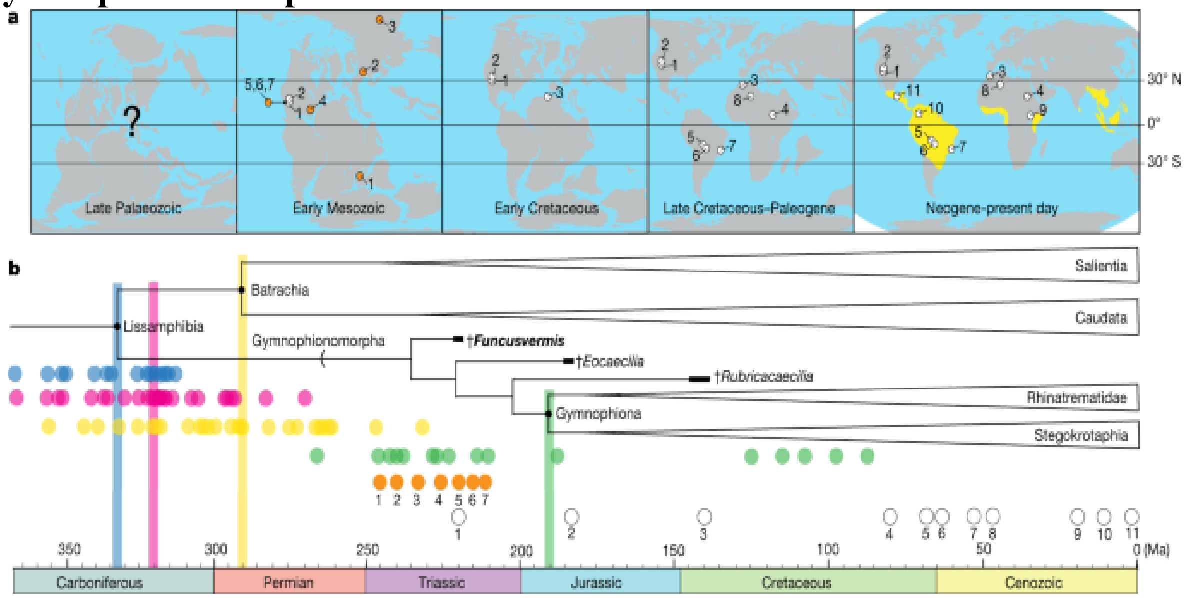
Given our phylogenetic results, the ecological habits of *Funcusvermis* may be transitional between terrestrial amphibamid dissorophoids and fossorial gymnophionans. The compound bones in the compact skull of fossorial gymnophionans are thought to withstand the forces associated with head-first burrowing[17], and at least some are present in *Funcusvermis* (for example, maxillopalatine). Small pits covering the lateral surfaces of the pseudodentary and maxillopalatine in *Funcusvermis* are also found in *Eocaecilia*[12], *Rubricacaecilia*[23] and gymnophionans[38]. External structure and internal microanatomy of these pits revealed by osteohistological sectioning of a *Funcusvermis* pseudodentary (PEFO 44432) show a marked resemblance to those of studied living caecilians (Extended Data Fig. 4), in which these pits act as anchor sites for collagen networks forming a tight skin-to-bone attachment and house glands that produce a lubricating mucus secretion, functions thought to aid in subterranean burrowing[39]. The dorsally

flattened neural arch of the *Funcusvermis* postatlantal pleurocentrum (PEFO 45810; Extended Data Fig. 3) resembles those of *Rubricacaecilia*[23], suggesting the acquisition of a tubular trunk, a feature crucial for underground locomotion in living caecilians[40]. These morphologies in *Funcusvermis* illustrate acquisition (by at least the Late Triassic) of some features that now facilitate fossoriality in living caecilians, later followed by acquisition of the DJCM and tentacular organ in *Eocaecilia*, and finally loss of the appendicular skeleton in gymnophionans.

Biogeography of early caecilians

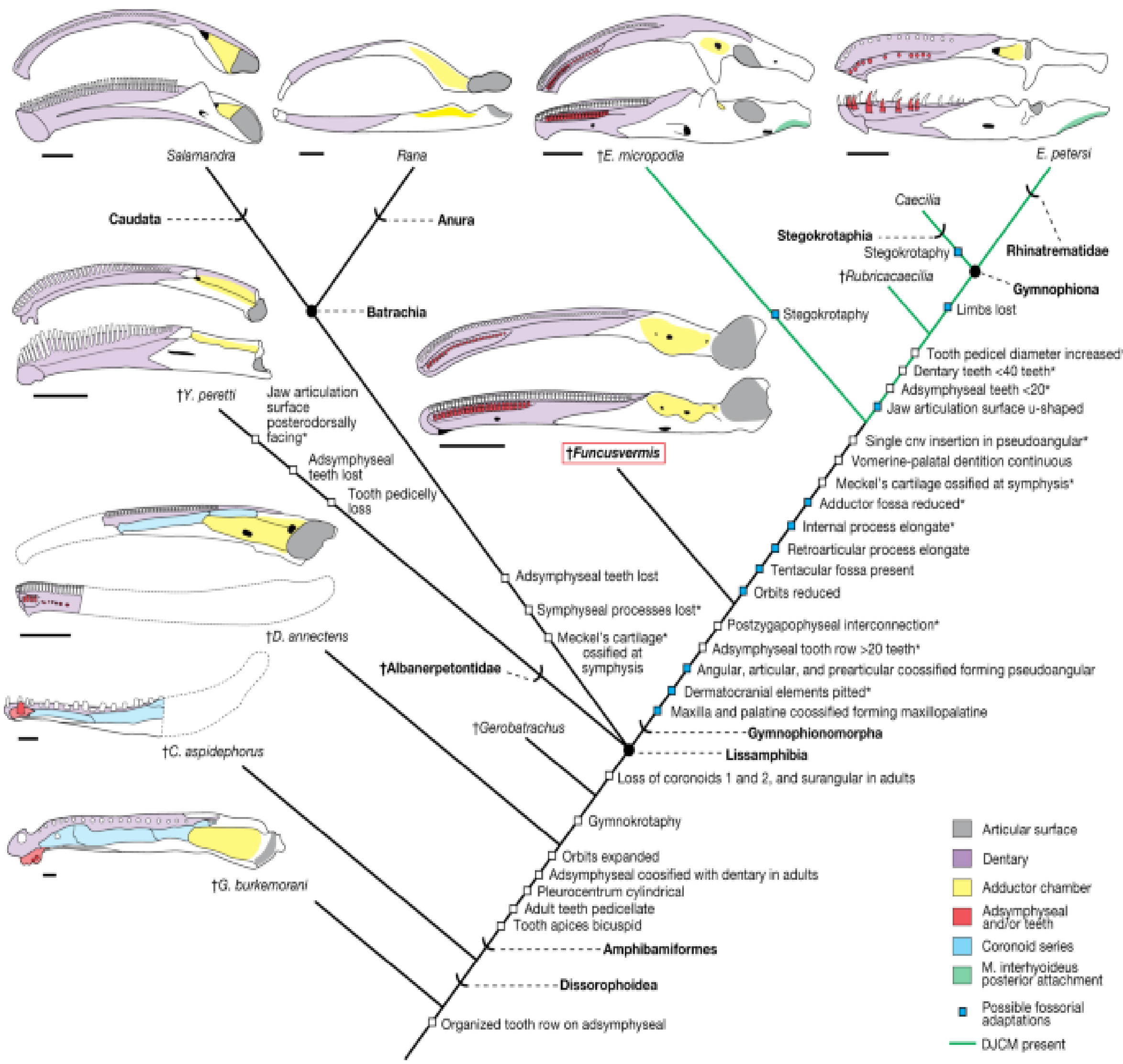
The spatiotemporal occurrence of *Funcusvermis* empirically establishes lissamphibian geographic origins on the Pangaean supercontinent before its fragmentation[20], and the similar palaeogeography of *Eocaecilia*[12] to *Funcusvermis* suggests the non-gymnophionan gymnophionomorph origin may lie in the early Mesozoic era of equatorial central Pangaea. The occurrence of *Rubricacaecilia* in the Early Cretaceous epoch of equatorial Gondwana may further support this hypothesis, suggesting non-gymnophionan gymnophionomorph distribution across both Laurasian and Gondwanan components of Pangaea in the early Mesozoic prior to its breakup[23]. The equatorial provenance of *Funcusvermis* adds to an exclusively equatorial pattern of gymnophionomorph distribution: all fossil occurrences fall between a minimum of approximately 16° N and 27° S (Fig. 2 and Supplementary Table 2), and living caecilians are restricted to equatorial latitudes[19] between 27° N and 34° S. The tropical distribution of extant gymnophionans is notably disjunct from non-gymnophionan gymnophionomorph fossil occurrences in present-day western North America and Morocco (Fig. 2). Drift of the North American and African plates during the Mesozoic[41] may explain the extirpation of gymnophionomorphs from these areas later in the Phanerozoic as these previously humid palaeotropical regions moved north into the arid subtropics. Concurrently, the northern drift of Gondwana into the palaeotropics may have expanded suitable terrestrial habitats, consistent with molecular evidence of an early Mesozoic Gondwanan origin of gymnophionans[20].

Fig. 2: Spatiotemporal history of Lissamphibia and Gymnophionomorpha.



a, Biogeographic history of Gymnophionomorpha and Triassic batrachians; yellow indicates living caecilian distribution. **b**, Time-calibrated topology of lissamphibian relationships showing major divergences (topology derived from refs. [6,23,38]). Estimated molecular divergence dates for major divergences are shown as blue circles (Gymnophionomopha–Batrachia divergence without *Gerobatrachus* calibration; Supplementary Table 4), pink circles (Gymnophionomopha–Batrachia divergence with *Gerobatrachus* calibration; Supplementary Table 5), yellow circles (Salientia–Caudata divergence; Supplementary Table 6) and green circles (Rhinatrematidae–Stegokrotaphia divergence; Supplementary Table 7); coloured vertical bars show the average for each set of divergence estimates. Numbered white and orange circles correspond to occurrences in Supplementary Tables 2 and 3, respectively. Crosses indicate extinct taxa.

Fig. 3: Evolutionary history of the lissamphibian mandibular ramus.



Squares denote important apomorphies (including non-mandibular features); apomorphies are optimized computationally unless followed by an asterisk, which denotes an apomorphy suggested by our results but lacking sufficient sampling to optimize computationally. Topology is derived from parsimony results (Extended Data Fig. 5); *Yaksha peretti*, *Salamandra* and *Rana* approximate conditions are found in taxa sampled in the analysis. Illustrations represent right mandibles in medial (bottom) and dorsal (top) views for *Doleserpeton annectens*[9], *Eocaecilia micropodia*[12] (Illustration adapted from ref. [12], with the permission of Museum of Comparative Zoology, Harvard University), *Epicrionops petersi*[12], *Funcusvermis gilmorei*, *Rana*, *Salamandra* and *Y. peretti*[28], excepting *Greererpeton burkemorani*[35] (dorsal only) and *Cacops aspidephorus*[37] (medial only). All

scale bars are 2 mm except for *G. burkemorani* (2 cm) and *C. aspidephorus* (2 cm). Brackets on the branches indicate stem groups, whereas circles indicate node groups. Crosses indicate extinct taxa.

The earliest batrachians hail from the Triassic of southern[2], equatorial[4,5] and northern[3,6] Pangaea (Supplementary Table 3), indicating extensive latitudinal dispersal by at least the Middle Triassic epoch; this pattern is further reflected in the subsequent batrachian fossil record and their extant distribution. Unlike in extant batrachians, evaporative water loss is found to be a critical physiological constraint in living caecilians, limiting their distribution to humid environments near the equator[19]. The contrasting spatiotemporal histories of batrachians and gymnophionomorphs suggest a divergence of physiological constraints linked to humidity prior to the Triassic; conserved physiological traits in these groups may explain subsequent patterns of dispersal reflected in present-day lissamphibian biogeography.

Timing of lissamphibian origins

Prior to the results of this study, the chronology of lissamphibian origins remained unresolved owing to the reliance of molecular clock estimates on different node minima derived from competing phylogenetic hypotheses that include extinct taxa[1]. *Funcusvermis* lends novel and strong support for a monophyletic origin of living amphibians within dissorophoid temnospondyls[30,42] (the 'classic' temnospondyl hypothesis), and thus the molecular clock estimates of caecilian–batrachian divergence using the temnospondyl hypothesis. Additionally, the recovery of *Gerobatrachus* as the sister taxon to Lissamphibia in our analysis suggests that taxon may not be a stem batrachian[7,8] and should be used with caution as a minimum age calibration for Lissamphibia. Molecular clock estimates using the temnospondyl hypothesis topology unconstrained by *Gerobatrachus* as the minimum age calibration of Lissamphibia may result in the most accurate estimates of the caecilian–batrachian divergence, and studies following these criteria show divergence time estimates ranging from the Late Devonian (367.0 Ma) to Middle Pennsylvanian (314.8 Ma) epochs, with a mean in the Middle Mississippian (333.5 Ma) and a median in the Late Mississippian (325.6 Ma) epoch (Fig. 2b and Supplementary Table 4). Our results refocus

the timeframe of lissamphibian origins to the Mississippian subperiod, older than previous estimates of a Pennsylvanian-Permian divergence based on calibrations using *Gerobatrachus* or *Amphibamus grandiceps*[1] and those considering Gymnophionomorpha as the sister group to the stereospondyl *Chinlestegophis*[1,14] (Fig. 2 and Supplementary Table 5).

Methods

New phylogenetic definition

Gymnophionomorpha Marjanović and Laurin 2008

Remarks. Gymnophionomorpha is defined here as the total group consisting of *Caecilia tentaculata* and all taxa that share a more recent common ancestor with it than with *Salamandra salamandra* Linnaeus, 1758, *Rana temporaria* Linnaeus, 1758, and *Albanerpeton inexpectatum*, Estes and Hoffstetter 1976. This newly proposed stem-based definition of Gymnophionomorpha is modified after that originally proposed[43].

Assignment of elements

Although all specimens assigned to *Funcusvermis* were found as isolated, dissociated elements, their assignment to a single gymnophionomorph taxon is supported by: (1) specimens bearing a suite of features present exclusively in gymnophionomorphs to the exclusion of all other tetrapods (Supplementary Information, section 2); (2) skeletal elements represented by multiple specimens (77 pseudodentaries and 8 pseudoangulars) where all are identical in morphology, varying only in size (Supplementary Information, sections 1 and 2); (3) the pseudodentary and pseudoangular bear complementary facets where they would overlap when in articulation (Extended Data Fig. 3).

Geological framework

The blue-coloured strata of the upper Blue Mesa Member of the Chinle Formation were deposited in a northwest-flowing fluviolacustrine system on

the western margin of central Pangaea at a palaeolatitude of 5° to 15° N in a humid monsoonal climate[26]. Detrital zircon U-Pb radiometric ages provide robust geochronologic constraints on the Chinle Formation, bracketing deposition of the upper Blue Mesa Member[24,25] to ~223–218 Ma (Extended Data Fig. 2). The gymnophionomorph fossils described herein were collected from the Thunderstorm Ridge locality (PFV 456) near the Puerco River in PEFO, Arizona, USA (Extended Data Fig. 2). The fossiliferous unit is a 15-cm-thick, poorly sorted siltstone horizon, bearing a dense concentration of carbonate nodules, angular intraformational clasts, micro- and macrovertebrate bones and coprolites. PFV 456 has yielded a diverse assemblage of vertebrates including chondrichthyans, actinopterygians, dipnoans, coelacanths, metoposaurids, salentians[5], drepanosauromorphs[44], lepidosauromorphs, archosauromorphs[45], pseudosuchian archosaurs[46], dinosauromorphs[47] and cynodonts[48]. The lack of abrasion and polishing and the exceptional three-dimensional preservation of extremely delicate microvertebrate bones indicates initial deposition in a low-energy setting, followed by brief reworking and redeposition in a channel avulsion event that incorporated angular intraformational clasts and carbonate nodules into the fossiliferous layer. This sedimentological evidence in combination with the presence of abundant spinicaudatan exoskeletons, unionid bivalve steinkerns and obligate-aquatic, amphibious and fully terrestrial vertebrates indicates initial deposition in a marginal lacustrine palaeoenvironment occupied by a diverse vertebrate community.

Collection and preparation methods

The hypodigm and all referred specimens were collected by screenwashing fossiliferous matrix from PFV 456 (9 out of 11 fossil gymnophionomorph occurrences were recovered using screenwashing; Supplementary Table 2). Blocks of matrix weighing approximately 1.8–3.2 kg were individually disaggregated in water and subsequently washed through a series of wire mesh screens with a minimum screen opening of 0.5 mm (no. 35 mesh). Dividing the fossiliferous concentrate from each block into smaller fractions in this way accelerated the process of picking. The resulting concentrate fractions were picked using a dissecting microscope resulting in the identification and separation of all *Funcusvermis* specimens. Importantly, through processing individual blocks of matrix, *Funcusvermis* elements that

fragmented into multiple parts during the screenwashing process could be re-associated after microscopic sorting. Elements found as multiple broken pieces were subsequently reassembled by adhering matching fractured surfaces using cyanoacrylate, typically a low viscosity PaleoBOND or Loctite brand. To facilitate rapidly and precisely adhering these miniscule fragments together we created a mechanism that combines aspects of a jeweler's block ball vice, and a hobbyist tool, sometimes called a third hand or helping hand. It combines a socket made of wood or closed cell polyethylene foam and a hemispherical wooden ball to create a pivot that can turn or tilt in all directions. This is topped with a small rectangle of wood with a small concave arch cut into it to provide a workspace. Insect pins are slid through channels in the wood filled with soft microcrystalline wax, which allows the pins freedom of movement, but the resistance needed to precisely position the fossil fragments. The fragments are temporarily adhered to the pin tips with more microcrystalline wax. Adhesive was applied to the joint between fragments as a microdroplet suspended on a single filament such as a cotton fibre and drawn into the joint via capillary action leaving a minimum of excess residue. Reassembly took place under a variety of Leica and Wild binocular microscopes, primarily MZ6, MZ12 and M8 models, varying in power from a maximum of ×40–×80 magnification.

To reveal the details of the pseudodentary dentition of *Funcusvermis*, matrix covering the dentition and other anatomy of PEFO 45800 was prepared through the following process. Melted cyclododecane (CDD) was poured into a shallow ceramic watch glass and allowed to harden. A small trench the size of the specimen was excavated, and the specimen was placed in the trench in the desired orientation. A Ukrainian *kistky* (a wax pen), was used to melt the CDD around the specimen and allowed it to adhere to and support the specimen. Then matrix was removed using a 1/32 inch (0.79375 mm) carbide-needle in a pin vice primarily under high magnification under a Leica MZ12 and MZ6 microscope. The point of the needle was ground to a superfine conical point at about 10°–15° parallel to the shaft and flattened briefly along one side to provide an edge to remove adhesives. Some of the softer clay particles were removed with a porcupine quill. When needed, the specimen was consolidated with a very dilute solution of polyvinyl butyral (Butvar B-76) in acetone; the solution was mixed by eye, applying a bit to another vertebrate bone fragment and looking for sheen upon drying. Any excess Butvar film was removed by abrasion with the porcupine quill. The

specimen was rotated in the CDD by trenching around the specimen until it was loose, shifting it, and then remelting the resulting CDD powder with the *kistky*. After all matrix was removed, the specimen was trenched out a final time and set aside in the fume hood to allow the CDD to sublimate.

Digital photography methods

Photographs of PEFO 45800 in Fig. 1 were acquired using a Leica MZ67 stereomicroscope and a Sony NEX-5T digital camera. Image stacking was conducted in Adobe Photoshop CC (https://www.adobe.com/products/photoshop.html).

Micro-computed tomographic scan methods

PEFO 44432, PEFO 45800, PEFO 45910, PEFO 46284, PEFO 46480 and PEFO 46481 were CT scanned with a Skyscan 1172 Microfocus X-radiographic Scanner at the Virginia Tech Institute for Critical Technology and Applied Science (ICTAS). PEFO 43891 was scanned with a Nikon XTH 225 ST High-Resolution X-ray Computed Tomography Scanner in the Shared Materials Instrumentation Facility at Duke University. Micro-computed scan parameters (resolution, source voltage, source current and scanning equipment type) for each scanned specimen included in Supplementary Table 8. Surface volume files (3D meshes) of specimens figured in Fig. 1 and Extended Data Figs. 3 and 4 are available for download under project 000382289 at Morphosource.org (https://www.morphosource.org/projects/000382289?locale=en).

3D segmentation methods

Scan datasets were processed using Dragonfly 2020.2 (http://www.theobjects.com/dragonfly) to produce 3D virtual reconstructions. PEFO 43891, PEFO 46284, and PEFO 46481 were segmented in Dragonfly 2020.2 to digitally remove matrix covering parts of the specimens.

Images of 3D surface meshes were produced using Meshlab 2021.07 (https://www.meshlab.net/).

Digital reconstruction methods

A composite reconstruction of a partial skull of *Funcusvermis* (Fig. 1) was produced using Meshmixer 3.5 (https://meshmixer.com). Digital 3D surface meshes representing the anterior (PEFO 43891; light pink in Fig. 1a–c) and posterior (PEFO 46284; dark pink in Fig. 1a–c) portions of a pseudodentary were scaled to the same dorsoventral height, and both specimens were overlapped to form a composite reconstruction of a complete pseudodentary. The pseudoangular (PEFO 46480) and maxillopalatine (PEFO 46481) were scaled to match the size of the reconstructed pseudodentary, and anatomically positioned relative to the pseudodentary to approximate their position in an articulated three-dimensional skull. A surface volume file (3D mesh) of the composite skull reconstruction is available for download under project 000382289 on Morphosource.org (https://www.morphosource.org/projects/000382289?locale=en).

Osteohistology methods

PEFO 44432 (right pseudodentary) was embedded in clear epoxy (Castolite AP), cut into 1 mm sections, and then ground to a ~100 μm thickness in the Virginia Tech Fossil Preparation Lab. Images of the histologically sectioned pseudodentary slide used in Extended Data Fig. 4 were acquired using a Sony NEX-5T digital camera mounted on a Nikon OPTIPHOT-POL Polarizing microscope. Fracturing of the specimen occurred during osteohistological preparation, causing fracture planes apparent in histological imaging (Extended Data Fig. 4).

Phylogenetic methods

See ‘Code availability’ to access and download phylogenetic matrix and analysis scripts.

Taxon sampling

Recent analyses recovered gymnophionomorphs at variable positions within Tetrapoda dependent on character and taxon sampling, including: (1) as ‘microsaur’ ‘lepospondyls’[49] (note that taxa formerly included in

‘Lepospondyli’ are now understood as polyphyletic[50]); (2) as stereospondyl temnospondyls forming the sister group to *C. jenkinsi*[14]; (3) as ‘microsaurian’ or aïstopod ‘lepospondyls’[51]; and [4] as amphibamiform dissorophoid temnospondyls forming the sister group to batrachians[6,13]. The matrix of Schoch et al. (2020), recently used to hypothesize the phylogenetic position of the stem salamander *Triassurus sixtelae* and the origin of lissamphibians, was selected to test the phylogenetic relationships of *F. gilmorei* given its comprehensive sampling of taxa proposed to be sister groups to Gymnophionomorpha including stem and crown amniotes, stereospondyl and dissorophoid temnospondyl amphibians, batrachians, gymnophionomorphs and albanerpetontids. *F. gilmorei* was coded into the modified Schoch et al. (2020) matrix, for a total of 63 sampled terminal taxa. See Supplementary Information, section 4 for discussion of taxon sampling.

Character sampling and scoring

Modifications to the Schoch et al. (2020) matrix are detailed in Supplementary Information, section 4 and include addition of new characters, modification of preexisting characters, exclusion of preexisting characters, and recodings of preexisting character states. *Funcusvermis* was coded for 29 characters in total based on currently known skeletal material (Supplementary Table 9). The final matrix includes 355 morphological characters (Full character list in Supplementary Information, section 8; see ‘Code availability’ to access and download phylogenetic matrix and analysis scripts).

Maximum parsimony and Bayesian analysis

All characters were equally weighted and unordered in both analyses following previous versions[6,14]. The character–taxon matrix was first analysed in the phylogenetic analysis software package TNT 1.5 (ref. [52]) using New Technology Search options with the following parameters: ratchet (1,000 iterations), sectoral search (1,000 rounds), tree fusing (100 rounds), and random additional sequence (1,000 replicates). A total of 71 most parsimonious trees of 1,468 steps each were recovered (consistency index = 0.287; retention index = 0.675). A strict consensus tree calculated from the most parsimonious trees is presented in Extended Data Fig. 5.

Bootstrap support values were obtained using TNT 1.5, and a strict consensus topology of trees produced via 1,000 bootstrap replicates resampled with replacement is presented in Extended Data Fig. 6. A Bayesian inference analysis of the character–taxon matrix was conducted in the phylogenetic software package MrBayes v.3.2.6 (ref. [53]) with the Mkv[54] model and gamma rate variation and the following parameters: four runs (six Markov chain Monte Carlo chains each), sampled every 1,000 generations, for 10 million generations with a relative burn-in of 0.25. Convergence of independent runs was assessed using Tracer v.1.76.1 (http://beast.bio.ed.ac.uk/Tracer). A consensus cladogram with mapped posterior probability values is presented in Extended Data Fig. 7.

Nomenclatural acts

The Life Science Identifiers (LSID) for the new genus and species are registered with Zoobank (http://zoobank.org) under the identifiers urn:lsid:zoobank.org:pub:A2A6C7AD-2077-413B-9004-2E841270A289.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The holotype, paratypes and referred specimens of *F. gilmorei* are catalogued and available for study to qualified researchers at PEFO. Computed tomographic scan data, including surface volume files (3D meshes) and raw CT data of *Funcusvermis* specimens mentioned in the main text and extended data figures (including the holotype, paratypes and referred specimens), as well as a surface volume file of the composite skull reconstruction of *Funcusvermis* are available for download under project 000382289 on Morphosource.org (https://www.morphosource.org/projects/000382289?locale=en).

Code availability

Code for TNT and MrBayes scripts used in the phylogenetic analyses conducted herein are available in Supplementary Information, section 8; the matrix is available for download under project 4166 on Morphobank.org (http://morphobank.org/permalink/?P4166).

References

1. Pardo, J. D., Lennie, K. & Anderson, J. S. Can we reliably calibrate deep nodes in the tetrapod tree? Case studies in deep tetrapod divergences. *Front. Genet.* **11**, 1159 (2020).

2. Rage, J.-C. & Roček, Z. Redescription of *Triadobatrachus massinoti* (Piveteau, 1936) an anuran amphibian from the early Triassic. *Palaeontographica A* **206**, 1–16 (1989).

3. Evans, S. E. & Borsuk-Białynicka, M. A stem-group frog from the Early Triassic of Poland. *Acta Palaeontol. Pol.* **43**, 573–580 (1998).

4. Heckert, A. B., Mitchell, J. S., Schneider, V. P. & Olsen, P. E. Diverse new microvertebrate assemblage from the Upper Triassic Cumnock Formation, Sanford Subbasin, North Carolina, USA. *J. Paleontol.* **86**, 368–390 (2012).

5. Stocker, M. R. et al. The earliest equatorial record of frogs from the Late Triassic of Arizona. *Biol. Lett.* **15**, 20180922 (2019).

6. Schoch, R. R., Werneburg, R. & Voigt, S. A Triassic stem-salamander from Kyrgyzstan and the origin of salamanders. *Proc. Natl Acad. Sci. USA* **117**, 11584–11588 (2020).

7. Anderson, J. S., Reisz, R. R., Scott, D., Fröbisch, N. B. & Sumida, S. S. A stem batrachian from the Early Permian of Texas and the origin of frogs and salamanders. *Nature* **453**, 515–518 (2008).

8. Anderson, J. S. Focal review: the origin(s) of modern amphibians. *Evol. Biol.* **35**, 231–247 (2008).

9. Sigurdsen, T. & Bolt, J. R. The Lower Permian amphibamid *Doleserpeton* (Temnospondyli: Dissorophoidea), the interrelationships of amphibamids, and the origin of modern amphibians. *J. Vertebr. Paleontol.* **30**, 1360–1377 (2010).

10. Schoch, R. R. The putative lissamphibian stem-group: phylogeny and evolution of the dissorophoid temnospondyls. *J. Paleontol.* **93**, 137–156 (2019).

11. Jenkins, P. A. & Walsh, D. M. An Early Jurassic caecilian with limbs. *Nature* **365**, 246–250 (1993).

12. Jenkins, F. A., Walsh, D. M. & Carroll, R. L. Anatomy of *Eocaecilia micropodia*, a limbed caecilian of the Early Jurassic. *Bull. Mus. Comp. Zool.* **158**, 285–365 (2007).

13. Maddin, H. C., Jenkins, F. A. Jr & Anderson, J. S. The braincase of *Eocaecilia micropodia* (Lissamphibia, Gymnophiona) and the origin of caecilians. *PLoS ONE* **7**, e50743 (2012).

14. Pardo, J. D., Small, B. J. & Huttenlocker, A. K. Stem caecilian from the Triassic of Colorado sheds light on the origins of Lissamphibia. *Proc. Natl Acad. Sci. USA* **114**, E5389–E5395 (2017).

15. Nussbaum, R. A. The evolution of a unique dual jaw-closing mechanism in caecilians: (Amphibia: Gymnophiona) and its bearing on caecilian ancestry. *J. Zool.* **199**, 545–554 (1983).

16. Kleinteich, T., Haas, A. & Summers, A. P. Caecilian jaw-closing mechanics: integrating two muscle systems. *J. R. Soc. Interface* **5**, 1491–1504 (2008).

17. Sherratt, E., Gower, D. J., Klingenberg, C. P. & Wilkinson, M. Evolution of cranial shape in caecilians (Amphibia: Gymnophiona). *Evol. Biol.* **41**, 528–545 (2014).

18. Schmidt, A. & Wake, M. H. Olfactory and vomeronasal systems of caecilians (Amphibia: Gymnophiona). *J. Morphol.* **205**, 255–268 (1990).

19. Pincheira-Donoso, D., Meiri, S., Jara, M., Olalla-Tárraga, M. Á. & Hodgson, D. J. Global patterns of body size evolution are driven by precipitation in legless amphibians. *Ecography* **42**, 1682–1690 (2019).

20. San Mauro, D., Vences, M., Alcobendas, M., Zardoya, R. & Meyer, A. Initial diversification of living amphibians predated the breakup of Pangaea. *Am. Nat.* **165**, 590–599 (2005).

21. Padian, K. & Sues, H.-D. in *Great Transformations in Vertebrate Evolution* (eds Dial, K. P., Shubin, N. & Brainerd, E. L.) 351–374 (Univ. Chicago Press, 2021).

22. Santos, R. O., Laurin, M. & Zaher, H. A review of the fossil record of caecilians (Lissamphibia: Gymnophionomorpha) with comments on its use to calibrate molecular timetrees. *Biol. J. Linn. Soc.* **131**, 737–755 (2020).

23. Evans, S. E. & Sigogneau-Russell, D. A stem-group caecilian (Lissamphibia: Gymnophiona) from the Lower Cretaceous of North Africa. *Palaeontology* **44**, 259–273 (2001).

24. Ramezani, J. et al. High-precision U-Pb zircon geochronology of the Late Triassic Chinle Formation, Petrified Forest National Park (Arizona, USA): temporal constraints on the early evolution of dinosaurs. *GSA Bull.* **123**, 2142–2159 (2011).

25. Rasmussen, C. et al. U-Pb zircon geochronology and depositional age models for the Upper Triassic Chinle Formation (Petrified Forest National Park, Arizona, USA): implications for Late Triassic paleoecological and paleoenvironmental change. *GSA Bull.* **133**, 539–558 (2021).

26. Nordt, L., Atchley, S. & Dworkin, S. Collapse of the Late Triassic megamonsoon in western equatorial Pangea, present-day American Southwest. *GSA Bull.* **127**, 1798–1815 (2015).

27. Martz, J. W. & Parker, W. G. in *Terrestrial Depositional Systems* (eds Zeigler, K. E. & Parker, W. G.) 39–125 (Elsevier, 2017).

28. Daza, J. D. et al. Enigmatic amphibians in mid-Cretaceous amber were chameleon-like ballistic feeders. *Science* **370**, 687–691 (2020).

29. Gardner, J. D. Monophyly and affinities of albanerpetontid amphibians (Temnospondyli; Lissamphibia). *Zool. J. Linn. Soc.* **131**, 309–352 (2001).

30. Bolt, J. R. Lissamphibian origins: possible protolissamphibian from the Lower Permian of Oklahoma. *Science* **166**, 888–891 (1969).

31. Gardner, J. D. & Averianov, A. O. Albanerpetontid amphibians from the Upper Cretaceous of Middle Asia. *Acta Palaeontol. Pol.* **43**, 453–476 (1998).

32. Carroll, R. L. The Palaeozoic ancestry of salamanders, frogs and caecilians. *Zool. J. Linn. Soc.* **150**, 1–140 (2007).

33. Müller, H., Oommen, O. V. & Bartsch, P. Skeletal development of the direct-developing caecilian *Gegeneophis ramaswamii* (Amphibia: Gymnophiona: Caeciliidae). *Zoomorphology* **124**, 171–188 (2005).

34. Ahlberg, P. E. & Clack, J. A. Lower jaws, lower tetrapods—a review based on the Devonian genus *Acanthostega*. *Earth Environ. Sci. Trans. R. Soc. Edinb.* **89**, 11–46 (1998).

35. Bolt, J. R. & Lombard, R. E. The mandible of the primitive tetrapod *Greererpeton*, and the early evolution of the tetrapod lower jaw. *J. Paleontol.* **75**, 1016–1042 (2001).

36. Shishkin, M. A. & Sulej, T. The Early Triassic temnospondyls of the Czatkowice 1 tetrapod assemblage. *Acta Palaeontol. Pol.* **65**, 31–77 (2009).

37. Anderson, J. S., Scott, D. & Reisz, R. R. The anatomy of the dermatocranium and mandible of *Cacops aspidephorus* Williston, 1910 (Temnospondyli: Dissorophidae), from the Lower Permian of Texas. *J. Vertebr. Paleontol.* **40**, e1776720 (2020).

38. Wilkinson, M., San Mauro, D., Sherratt, E. & Gower, D. J. A nine-family classification of caecilians (Amphibia: Gymnophiona). *Zootaxa* **2874**, 41–64 (2011).

39. Jared, C. et al. Skin gland concentrations adapted to different evolutionary pressures in the head and posterior regions of the caecilian *Siphonops annulatus*. *Sci. Rep.* **8**, 3576 (2018).

40. O'Reilly, J. C., Ritter, D. A. & Carrier, D. R. Hydrostatic locomotion in a limbless tetrapod. *Nature* **386**, 269–272 (1997).

41. Muttoni, G. & Kent, D. V. Jurassic monster polar shift confirmed by sequential paleopoles from Adria, promontory of Africa. *J. Geophys. Res.* **124**, 3288–3306 (2019).

42. Parsons, T. S. & Williams, E. E. The relationships of the modern Amphibia: a re-examination. *Q. Rev. Biol.* **38**, 26–53 (1963).

43. Marjanović, D. & Laurin, M. A reevaluation of the evidence supporting an unorthodox hypothesis on the origin of extant amphibians. *Contrib. Zool.* **77**, 149–199 (2008).

44. Jenkins, X. A. et al. Using manual ungual morphology to predict substrate use in the Drepanosauromorpha and the description of a new species. *J. Vertebr. Paleontol.* **40**, e1810058 (2020).

45. Kligman, B. T., Marsh, A. D., Nesbitt, S. J., Parker, W. G. & Stocker, M. R. New trilophosaurid species demonstrates a decline in allokotosaur diversity across the Adamanian–Revueltian boundary in the Late Triassic of western North America. *Palaeodiversity* **13**, 25–37 (2020).

46. Marsh, A. D., Smith, M. E., Parker, W. G., Irmis, R. B. & Kligman, B. T. Skeletal anatomy of *Acaenasuchus geoffreyi* Long and Murry, 1995 (Archosauria: Pseudosuchia) and its implications for the origin of the aetosaurian carapace. *J. Vertebr. Paleontol.* **40**, e1794885 (2020).

47. Marsh, A. D. & Parker, W. G. New dinosauromorph specimens from Petrified Forest National Park and a global biostratigraphic review of

Triassic dinosauromorph body fossils. *PaleoBios* https://doi.org/10.5070/P9371050859 (2020).

48. Kligman, B. T., Marsh, A. D., Sues, H.-D. & Sidor, C. A. A new non-mammalian eucynodont from the Chinle Formation (Triassic: Norian), and implications for the early Mesozoic equatorial cynodont record. *Biol. Lett.* **16**, 20200631 (2020).

49. Huttenlocker, A. K., Pardo, J. D., Small, B. J. & Anderson, J. S. Cranial morphology of recumbirostrans (Lepospondyli) from the Permian of Kansas and Nebraska, and early morphological evolution inferred by micro-computed tomography. *J. Vertebr. Paleontol.* **33**, 540–552 (2013).

50. Pardo, J. D., Szostakiwskyj, M., Ahlberg, P. E. & Anderson, J. S. Hidden morphological diversity among early tetrapods. *Nature* **546**, 642–645 (2017).

51. Marjanović, D. & Laurin, M. Phylogeny of Paleozoic limbed vertebrates reassessed through revision and expansion of the largest published relevant data matrix. *PeerJ* **6**, e5565 (2019).

52. Goloboff, P. A. & Catalano, S. A. TNT version 1.5, including a full implementation of phylogenetic morphometrics. *Cladistics* **32**, 221–238 (2016).

53. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).

54. Lewis, P. O. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* **50**, 913–925 (2001).

55. Eltink, E., Schoch, R. R. & Langer, M. C. Interrelationships, palaeobiogeography and early evolution of Stereospondylomorpha (Tetrapoda: Temnospondyli). *J. Iber. Geol.* **45**, 251–267 (2019).

56. Bystrow, A. *Dvinosaurus* als neotenische Form der Stegocephalen. *Acta Zool.* **19**, 209–295 (1938).

57. Dutuit, J.-M. Introduction à l'étude paléontologique du Trias continental Marocain. Description des premiers stegocephales recueillis dans le couloir d'Argana (Atlas Occidental). *Mémoires du Muséum National d'Histoire* **36**, 1–253 (1976).

58. Dias, E. V., Dias-da-Silva, S. & Schultz, C. L. A new short-snouted rhinesuchid from the Permian of southern Brazil. *Revista Brasileira de Paleontologia* **23**, 98–122 (2020).

59. Damiani, R. J. & Kitching, J. W. A new brachyopid temnospondyl from the *Cynognathus* Assemblage Zone, Upper Beaufort Group, South Africa. *J. Vertebr. Paleontol.* **23**, 67–78 (2003).

60. Schoch, R. R. & Witzmann, F. Cranial morphology of the plagiosaurid *Gerrothorax pulcherrimus* as an extreme example of evolutionary stasis. *Lethaia* **45**, 371–385 (2012).

61. Schoch, R. R. Studies on braincases of early tetrapods: Structure, morphological diversity, and phylogeny-1 Trimerorhacis and other prmitive temnospondyls. *Neues Jahrbuch für Geologie und Paläontologie-Abhandlungen* **213**, 233–259 (1999).

62. Ruta, M. & Bolt, J. R. The brachyopoid *Hadrokkosaurus bradyi* from the early Middle Triassic of Arizona, and a phylogenetic analysis of lower jaw characters in temnospondyl amphibians. *Acta Palaeontol. Pol.* **53**, 579–592 (2008).

63. Bystrow, A. & Efremov, J. *Benthosuchus sushkini* Efr.—a labyrinthodont from the Eotriassic of Sharzhenga River. *Trudy Paleontol. Inst.* **10**, 1–152 (1940).

64. Warren, A. Karoo tupilakosaurid: a relict from Gondwana. *Earth Environ. Sci. Trans. R. Soc. Edinb.* **89**, 145–160 (1998).

65. Holmes, R. B., Carroll, R. L. & Reisz, R. R. The first articulated skeleton of *Dendrerpeton acadianum* (Temnospondyli, Dendrerpetontidae) from the Lower Pennsylvanian locality of Joggins,

Nova Scotia, and a review of its relationships. *J. Vertebr. Paleontol.* **18**, 64–79 (1998).

66. Steyer, J. S. The first articulated trematosaur ‘amphibian’ from the Lower Triassic of Madagascar: implications for the phylogeny of the group. *Palaeontol.* **45**, 771–793 (2002).

67. Englehorn, J., Small, B. J. & Huttenlocker, A. A redescription of *Acroplous vorax* (Temnospondyli: Dvinosauria) based on new specimens from the Early Permian of Nebraska and Kansas, USA. *J. Vertebr. Paleontol.* **28**, 291–305 (2008).

68. Warren, A. *Laidleria* uncovered: a redescription of *Laidleria gracilis* Kitching (1957), a temnospondyl from the Cynognathus Zone of South Africa. *Zool. J. Linn. Soc.* **122**, 167–185 (1998).

69. Bolt, J. R. & Chatterjee, S. A new temnospondyl amphibian from the Late Triassic of Texas. *J. Paleontol.* **74**, 670–683 (2000).

70. Milner, A. & Sequeira, S. The temnospondyl amphibians from the Viséan of east Kirkton, West Lothian, Scotland. *Earth Environ. Sci. Trans. R. Soc. Edinb.* **84**, 331–361 (1993).

71. Schoch, R. R. & Milner, A. R. *Encyclopedia of Paleoherpetology, Part 3A. Temnospondyli* (Verlag Dr. Friedrich Pfeil, 2014).

72. Damiani, R., Schoch, R. R., Hellrung, H., Werneburg, R. & Gastou, S. The plagiosaurid temnospondyl *Plagiosuchus pustuliferus* (Amphibia: Temnospondyli) from the Middle Triassic of Germany: anatomy and functional morphology of the skull. *Zool. J. Linn. Soc.* **155**, 348–373 (2009).

73. Chernin, S. A new brachyopid, *Batrachosuchus concordi* sp. nov. from the Upper Luangwa Valley, Zambia with a redescription of *Batrachosuchus browni* Broom, 1903. *Palaeontol. Afr.* **20**, 87–109 (1977).

74. Sulej, T. Osteology, variability, and evolution of *Metoposaurus*, a temnospondyl from the Late Triassic of Poland. *Acta Palaeontol. Pol.*

64, 29–139 (2007).

Acknowledgements

We thank C. Beightol, W. Reyes, E. Patellos, X. Jenkins, E. Smith and members of the 2019 Virginia Tech Paleobiology Research Group field team for their fieldwork efforts. D. Wagner, D. Boudreau, P. Varela, L. McWhinney, N. Pezzoni, N. Green, C. Heltzel, C. Sheehy and Z. Lester conducted microscopic picking resulting in the collection of many specimens. C. Lash, D. Wagner, D. Boudreau and P. Varela provided logistical support in the PEFO laboratory and collections. R. Mueller and J. Socha facilitated µCT scanner access at Virginia Tech. J. Gladman facilitated µCT scanner access at the Shared Materials Instrumentation Facility at Duke University. J. Gillette and D. Gillette (MNA) provided access to *Eocaecilia* specimens. J. Martz illustrated the stratigraphy of PEFO in Extended Data Fig. 2. B. Creisler assisted in formulating the name of the new taxon. T. Young facilitated funding from the Petrified Forest Museum Association for the illustration of *F. gilmorei* by A. Atuchin. This study was supported by the David R. Wones Geological Scholarship (to B.T.K.), the USA National Park Service (PMIS 209814 to B.T.K., W.G.P. and A.D.M.), the Virginia Tech Department of Geosciences (to B.T.K., M.R.S. and S.J.N.), the National Science Foundation (DEB1655609 to M.R.S.; EAR480383 to S.J.N., M.R.S. and B.T.K.; and DPP1947094 to C. Sidor, supporting B.M.G.), the David B. Jones Foundation (to M.R.S. and S.J.N.) and the Petrified Forest Museum Association (to B.T.K.). We thank the Willi Henning Society for free access to TNT software. This is Petrified Forest National Park Contribution no. 87. Views expressed herein are those of the authors and do not represent the views of the United States Government. Finally, we thank James 'Diamond' Williams and the Ohio Players for their permission to name the genus after their song *Funky Worm* from the 1972 album *Pleasure*.

Author information

Authors and Affiliations

1. Department of Resource Management and Science, Petrified Forest National Park, Petrified Forest, AZ, USA

   Ben T. Kligman, Adam D. Marsh, Matthew E. Smith & William G. Parker

2. Department of Geosciences, Virginia Tech, Blacksburg, VA, USA

   Ben T. Kligman, Sterling J. Nesbitt & Michelle R. Stocker

3. Burke Museum and Department of Biology, University of Washington, Seattle, WA, USA

   Bryan M. Gee

Contributions

B.T.K., A.D.M., W.G.P. and M.R.S. conceived the project. B.T.K., A.D.M., W.G.P., M.R.S. and S.J.N. conducted fieldwork collecting fossiliferous sediment. M.E.S., B.T.K. and A.D.M. conducted laboratory work identifying and physically preparing specimens. B.T.K. and M.R.S. sampled µCT data. B.T.K. conducted µCT digital processing and preparation. S.J.N. and B.T.K. conducted histological sampling. B.M.G. led revisions of the phylogenetic dataset with assistance from B.T.K. B.M.G. and B.T.K. coded the phylogenetic matrix. B.T.K. and B.M.G. conducted phylogenetic analyses. B.T.K., A.D.M. and B.M.G. prepared figures. B.T.K. wrote the manuscript, with contributions and editing from all authors.

Corresponding authors

Correspondence to Ben T. Kligman, Bryan M. Gee, Adam D. Marsh, Sterling J. Nesbitt, Matthew E. Smith, William G. Parker or Michelle R. Stocker.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Jason Anderson, Rainer Schoch and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Life reconstruction of *Funcusvermis gilmorei* (lower) and *Acaenasuchus geoffreyi* (upper) in a paleoenvironmental reconstruction of PFV 456.

Illustration by Andrey Atuchin.

Extended Data Fig. 2 Geographic and stratigraphic context of PFV 456 (Thunderstorm Ridge), the type locality of *Funcusvermis gilmorei* in the upper Blue Mesa Member of the Chinle Formation (Late Triassic) at Petrified Forest National Park, AZ, USA.

U-Pb radioisotopic ages derived from refs. [24,25].

Extended Data Fig. 3 Anatomy of *Funcusvermis gilmorei* based on reconstructed 3D surface meshes from segmented micro-computed tomographic scan data.

a–**c**, Referred left pseudoangular (PEFO 46480) in ventral, lateral, and posterior views. **e**–**i**, Holotype right pseudodentary (PEFO 43891) in medial, anteromedial, anterior, ventral, dorsal, and anterolateral views. **j**–**l**, Paratype right pseudodentary (PEFO 46284) in medial, dorsal, and lateral views. **m**–**o**, Paratype right pseudodentary (PEFO 45800) in dorsal, medial, and lateral views. **p**, Paratype right pseudodentary (PEFO 46284) and referred left pseudoangular (PEFO 46480) with yellow arrows showing matching facet surfaces and neurovascular canals. **q**–**u**, Referred postatlantal vertebra (PEFO 45810) in dorsal, ventral, right lateral, anterior, and posterior views. **v**–**y**, Referred right femur (PEFO 43811) in dorsal, ventrolateral, lateral, and proximal views. abcnV, alveolar branch cranial nerve V; adtr, adsymphyseal tooth row; aeecnV, anterior exit external branch of cranial nerve V; c, centrum; cnVII, cranial nerve VII insertion; dp, diapophysis; dpaf, dorsal pseudoangular facet; dpdf, dorsal pseudodentary facet; dtr, dentary tooth row; ebcnV, external branch cranial nerve V; fh, femoral head; lasdg, labial subdental groove; ld, lateral depression; lecnV, lateral exit cranial nerve V; lisdg, lingual subdental groove; mf, medial fossa; mkv; midventral keel; ms, mandibular symphysis; mt, medial trochanter; na, neural arch; nc, neural canal; p, pit; paaspd, pseudoangular attachment surface of the pseudodentary; pap, posterior pseudoangular process; pdaspa, pseudodentary attachment surface of the pseudoangular; peecnV, posterior exit external branch cranial nerve V; poz, postzygapophysis; prez, prezygapophysis; ptadtr, posterior terminus adsymphyseal tooth row; ptdtr, posterior terminus dentary tooth row; sf, symphyseal foramen; sp, symphyseal prongs; vpaf, ventral pseudoangular facet; vpdf, ventral pseudodentary facet. Arrows point in anterior direction. Scale bars equal 1 mm.

Extended Data Fig. 4 Cross sectional views of *Funcusvermis gilmorei* dentigerous elements from micro-computed tomographic scan data and osteohistological sectioning.

a–**d**, Coronal cross sections of referred left maxillopalatine (PEFO 46481). **e**–**k**, Coronal cross sections of holotype right pseudodentary (PEFO 43891). **l**–**n**, Coronal cross sections of paratype right pseudodentary (PEFO 46284). **o**–**p**, Microanatomy from osteohistological section of paratype right pseudodentary (PEFO 44432) in coronal view. **r**, Referred left maxillopalatine (PEFO 46481) in lateral view. Paratype right pseudodentary

(PEFO 44432) in lateral (**s**) and medial (**t**) views. abcnV, alveolar branch cranial nerve V; dpdf, dorsal pseudodentary facet; ebcnV, external branch cranial nerve V; frac, fractures; fr, facial ramus; ig, interdental groove; laam, labial alveolar margin; lapw, labial pedicel wall; lasdg, labial subdental groove; liam, lingal alveolar margin; lipw, lingual pedicel wall; lisdg, lingual subdental groove; oc, osteocyte lacunae; om, orbital margin; p, pit; pc, pulp cavity; peecnV, posterior exit external branch cranial nerve V; sf, symphyseal foramen; vpdf, ventral pseudodentary facet. Red lines indicate locations of cross sections. Yellow outlines denote margins of canals connected to external pits. Scale bars equal 1 mm unless otherwise noted.

Extended Data Fig. 5 Phylogenetic tree from parsimony analysis.

Strict consensus tree of 71 most parsimonious trees (1,468 steps each) from TNT parsimony analysis (Methods). CI = 0.287, RI = 0.675.

Extended Data Fig. 6 Phylogenetic tree from parsimony analysis showing bootstrap support.

Strict consensus topology of trees produced via 1,000 bootstrap replicates resampled with replacement (Methods). Node values indicate bootstrap support.

Extended Data Fig. 7 Topology from Bayesian inference phylogenetic analysis.

Majority-rule consensus tree from Bayesian inference analysis (Methods). Node labels show posterior probability values.

Extended Data Fig. 8 Phylogenetic distribution and comparative morphology of a single row of coronoid teeth in Temnospondyli.

a, Phylogenetic topology (derived from refs. [14,55]) with presence of a semi-continuous row of teeth across all three coronoids (See Supplementary

Information 3 for further discussion of this feature). **b**–**k**, Comparative silhouettes of lower jaws in lingual (**b**–**g**), and dorsal (**h**–**m**), views. **b**, The dvinosaur *Dvinosaurus primus*[56]. **c**, The trematosaur *Almasaurus habbazi*[57]. **d**, The rhinesuchid *Rastosuchus hammeri*[58]. **e**, The brachyopid *Vanastega plurimidens*[59]. **f**, The stereospondyl *Chinlestegophis jenkinsi*[14]. **g**, The plagiosaurid *Gerrothorax pulcherrimus*[60] (Illustration adapted from ref. [60] under CC BY 3.0 license). **h**, The capitosaur *Mastodonsaurus giganteus*[61]. **i**, The brachyopid *Hadrokkosaurus bradyi*[62]. **j**, The trematosaur *Benthosuchus sushkini*[63]. **k**, *Chinlestegophis jenkinsi*[14]. Arrows indicate anterior direction.

Extended Data Fig. 9 Phylogenetic distribution and comparative morphology of a lateral exposure of the palatine (LEP) and the loss of the lacrimal in Temnospondyli.

a, Phylogenetic topology (derived from refs. [14,55]) with variable presence of the lacrimal and the LEP (See Supplementary Information 3 for further discussion of this feature). Comparative silhouettes of skulls in dorsal **b**–**h** and lateral **i**–**k** views; the prefrontal is coloured yellow here as it is the main element to positionally compensate for an absent lacrimal. **b**, The dvinosaur *Thabanchuia oomie*[64] (Illustration reproduced with permission from Cambridge University Press from ref. [64]). **c**, The amphibamid *Doleserpeton annectens*[9] (Illustration adapted from ref. [9] under copyright © Society of Vertebrate Paleontology, www.vertpaleo.org, reprinted by permission of Informa UK Limited, trading as Taylor & Francis Group, www.tandfonline.com on behalf of Society of Vertebrate Paleontology, www.vertpaleo.org. **d**, The 'dendrerpetid' *Dendrerpeton helogenes*[65], which represents the plesiomorphic condition (Illustration adapted from ref. [65]. under copyright © Society of Vertebrate Paleontology, www.vertpaleo.org, reprinted by permission of Informa UK Limited, trading as Taylor & Francis Group, www.tandfonline.com on behalf of Society of Vertebrate Paleontology, www.vertpaleo.org). **e**, The trematosaur *Wantzosaurus elongatus*[66]. **f**, The dvinosaur *Acroplous vorax*[67] (Illustration adapted from ref. [67] under copyright © Society of Vertebrate Paleontology, www.vertpaleo.org, reprinted by permission of Informa UK Limited, trading as Taylor & Francis Group, www.tandfonline.com on behalf of Society of Vertebrate Paleontology, www.vertpaleo.org). **g**, the rhytidosteid *Laidleria*

gracilis[68] (Illustration adapted from ref. [68]. by permission of the Zoological Journal of the Linnean Society). **h**, *Chinlestegophis jenkinsi*[14]. **i**, *Acroplous vorax*[67] in lateral view. **j**, *Rileymillerus cosgriffi*[69] in lateral view. **k**, *Chinlestegophis jenkinsi*[14] in lateral view. The prefrontal is also coloured yellow here as it is the main element to positionally compensate for an absent lacrimal. Silhouettes are not to scale.

Extended Data Fig. 10 Phylogenetic distribution and comparative morphology of a broad cultriform process and a posteriorly situated occiput in Temnospondyli.

a, Phylogenetic topology (derived from refs. [14,55]) with presence of a broad cultriform process and a posteriorly situated occiput (See Supplementary Information 3 for further discussion of this feature). **b**–**j**, Comparative silhouettes of skulls in lateral **i**–**k**, and ventral **b**–**h**, views. **b**, The dvinosaur *Dvinosaurus primus*[56]. **c**, The ‘dendrerpetid’ *Balanerepeton woodi*[70,71] (Illustration adapted from ref. [70]. with permission from Cambridge University Press), which represents the plesiomorphic condition. **d**, The plagiosaurid *Plagiosuchus pustuliferus*[72] (Illustration adapted from ref. [72]. by permission of the Zoological Journal of the Linnean Society). **e**, The brachyopid *Batrachosuchus browni*[73]. **f**, The metoposaurid *Metoposaurus krasiejowensis*[74]. **g**, *Chinlestegophis jenkinsi*[14]. **h**, *Batrachosuchus browni*[73]. **i**, The dvinosaur *Thabanchuia oomie*[64] (Illustration reproduced with permission from Cambridge University Press from ref. [64]). **j**, *Chinlestegophis jenkinsi*[14]. Silhouettes are not to scale.

Supplementary information

Supplementary Information

This supplementary information file contains the following sections: 1. Complete list of material assigned to *F. gilmorei*. 2. Expanded description of *F. gilmorei*. 3. Comparisons to stereospondyl temnospondyls and *C. jenkinsi*. 4. Revisions and additions to phylogenetic dataset. 5. Supplementary Tables

1–9. 6. Institutional abbreviations. 7. Supplementary references. 8. Character list and phylogenetic datasets.

[Reporting Summary]

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Kligman, B.T., Gee, B.M., Marsh, A.D. *et al.* Triassic stem caecilian supports dissorophoid origin of living amphibians. *Nature* **614**, 102–107 (2023). https://doi.org/10.1038/s41586-022-05646-5

- Received: 12 January 2022
- Accepted: 12 December 2022
- Published: 25 January 2023
- Issue Date: 02 February 2023

- DOI: https://doi.org/10.1038/s41586-022-05646-5

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

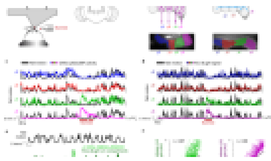
Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

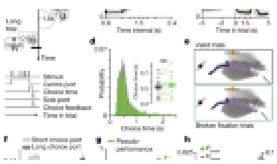
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05646-5



Context-dependent representations of movement in Drosophila dopaminergic reinforcement pathways

25 October 2021

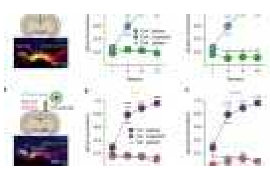
Aryeh Zolin, Raphael Cohn, … Vanessa Ruta



Action suppression reveals opponent parallel control via striatal circuits

06 July 2022

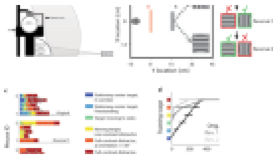
Bruno F. Cruz, Gonçalo Guiomar, … Joseph J. Paton



Dopamine neurons create Pavlovian conditioned stimuli with circuit-defined motivational properties

23 July 2018

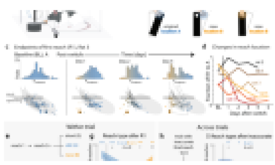
Benjamin T. Saunders, Jocelyn M. Richard, … Patricia H. Janak



[The Virtual-Environment-Foraging Task enables rapid training and single-trial metrics of rule acquisition and reversal in head-fixed mice]

18 March 2019

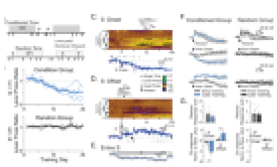
Martha N. Havenith, Peter M. Zijderveld, … Jeffrey C. Glennon



[Transition from predictable to variable motor cortex and striatal ensemble patterning during behavioral exploration]

04 May 2022

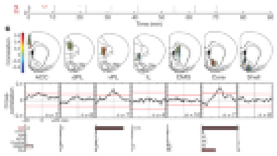
Sravani Kondapavulur, Stefan M. Lemke, … Karunesh Ganguly



[Dopamine encodes real-time reward availability and transitions between reward availability states on different timescales]

01 July 2022

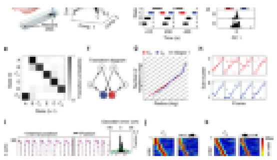
Abigail Kalmbach, Vanessa Winiger, … Eleanor H. Simpson



[Dissociable dopamine dynamics for learning and motivation]

22 May 2019

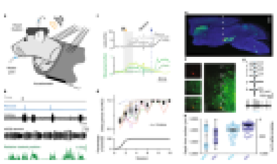
Ali Mohebi, Jeffrey R. Pettibone, … Joshua D. Berke



[Revealing neural correlates of behavior without behavioral measurements]

18 October 2019

Alon Rubin, Liron Sheintuch, … Yaniv Ziv



[The timing of action determines reward prediction signals in identified midbrain dopamine neurons]

15 October 2018

Luke T. Coddington & Joshua T. Dudman

- Article
- Open Access
- Published: 18 January 2023

Spontaneous behaviour is structured by reinforcement without explicit reward

- Jeffrey E. Markowitz,
- Winthrop F. Gillis,
- Maya Jay,
- Jeffrey Wood,
- Ryley W. Harris,
- Robert Cieszkowski,
- Rebecca Scott,
- David Brann,
- Dorothy Koveal,
- Tomasz Kula,
- Caleb Weinreb,
- Mohammed Abdal Monium Osman,
- Sandra Romero Pinto,
- Naoshige Uchida,
- Scott W. Linderman,
- Bernardo L. Sabatini &
- …
- Sandeep Robert Datta

Show authors

Nature **volume 614**, pages 108–117 (2023)

- 25k Accesses
- 2 Citations
- 254 Altmetric
- Metrics details

Abstract

Spontaneous animal behaviour is built from action modules that are concatenated by the brain into sequences[1,2]. However, the neural mechanisms that guide the composition of naturalistic, self-motivated behaviour remain unknown. Here we show that dopamine systematically fluctuates in the dorsolateral striatum (DLS) as mice spontaneously express sub-second behavioural modules, despite the absence of task structure, sensory cues or exogenous reward. Photometric recordings and calibrated closed-loop optogenetic manipulations during open field behaviour demonstrate that DLS dopamine fluctuations increase sequence variation over seconds, reinforce the use of associated behavioural modules over minutes, and modulate the vigour with which modules are expressed, without directly influencing movement initiation or moment-to-moment kinematics. Although the reinforcing effects of optogenetic DLS dopamine manipulations vary across behavioural modules and individual mice, these differences are well predicted by observed variation in the relationships between endogenous dopamine and module use. Consistent with the possibility that DLS dopamine fluctuations act as a teaching signal, mice build sequences during exploration as if to maximize dopamine. Together, these findings suggest a model in which the same circuits and computations that govern action choices in structured tasks have a key role in sculpting the content of unconstrained, high-dimensional, spontaneous behaviour.

Main

Spontaneous behaviour exhibits structure. Ethologists have long argued that the self-motivated behaviour of animals in the wild is flexibly built from modular components that are linked together over time in a predictable yet probabilistic manner[1]. Many well-studied laboratory behaviours—including chemotaxis, grooming, prey seeking, courtship, birdsong and exploratory locomotion—are similarly characterized by modularity and predictability[2,3,4,5]. However, it remains unclear how the brain regulates the expression of individual behavioural modules at any given moment, or how it dynamically composes these modules into the fluid behaviours observed

when animals act of their own volition in the absence of experimental restraint, task structure or exogenous reward.

Given that the loss of dopaminergic neurons from the substantia nigra pars compacta (SNc) causes diffuse deficits in action initiation and sequencing, it is likely that the neuromodulator dopamine influences the architecture of spontaneous behaviour[6,7,8]. Yet we know little about the precise relationship between dopamine and behaviour when animals freely explore an environment. Although dopamine is thought to motivate spontaneous behaviour and to influence the vigour with which actions are expressed, evidence is mixed as to whether phasic dopamine transients are permissive or causal for movements, whether dopamine rises or falls when animals initiate a movement, and whether dopamine fluctuations specify movement kinematics in freely behaving animals[6,9,10,11,12,13,14,15,16,17,18,19]. By contrast, during structured tasks in which animals seek explicit and often cued rewards, phasic dopamine clearly conveys information related to reward and reward-prediction errors, reinforces reward-associated actions, and influences the choices made between alternative actions[20,21,22,23,24,25].

Dopamine may have distinct roles during spontaneous and task-structured behaviours, given the many ways in which they differ; for example, spontaneous behaviours generally exhibit a greater variety of expressed behavioural modules, include more complex behavioural sequences, and tend to emphasize self-initiated movements associated with active sensing[2,4,26]. Nevertheless, both spontaneous behaviour and structured tasks demand that animals choose actions on an ongoing basis from a distribution of possibilities, suggesting that dopamine may influence the continuous assembly of naturalistic sequences through mechanisms similar to those used to support goal-driven action selection in response to rewards.

To test this hypothesis, here we characterize mouse spontaneous behaviour using motion sequencing (MoSeq)—which uses 3D imaging and unsupervised machine learning to atomize behaviour into sub-second modules referred to as ‘syllables’—while simultaneously assessing and manipulating dopamine transients in DLS, a region of the basal ganglia implicated in the composition of natural behaviours[27,28,29]. As SNc dopaminergic neurons acutely influence the population activity of DLS

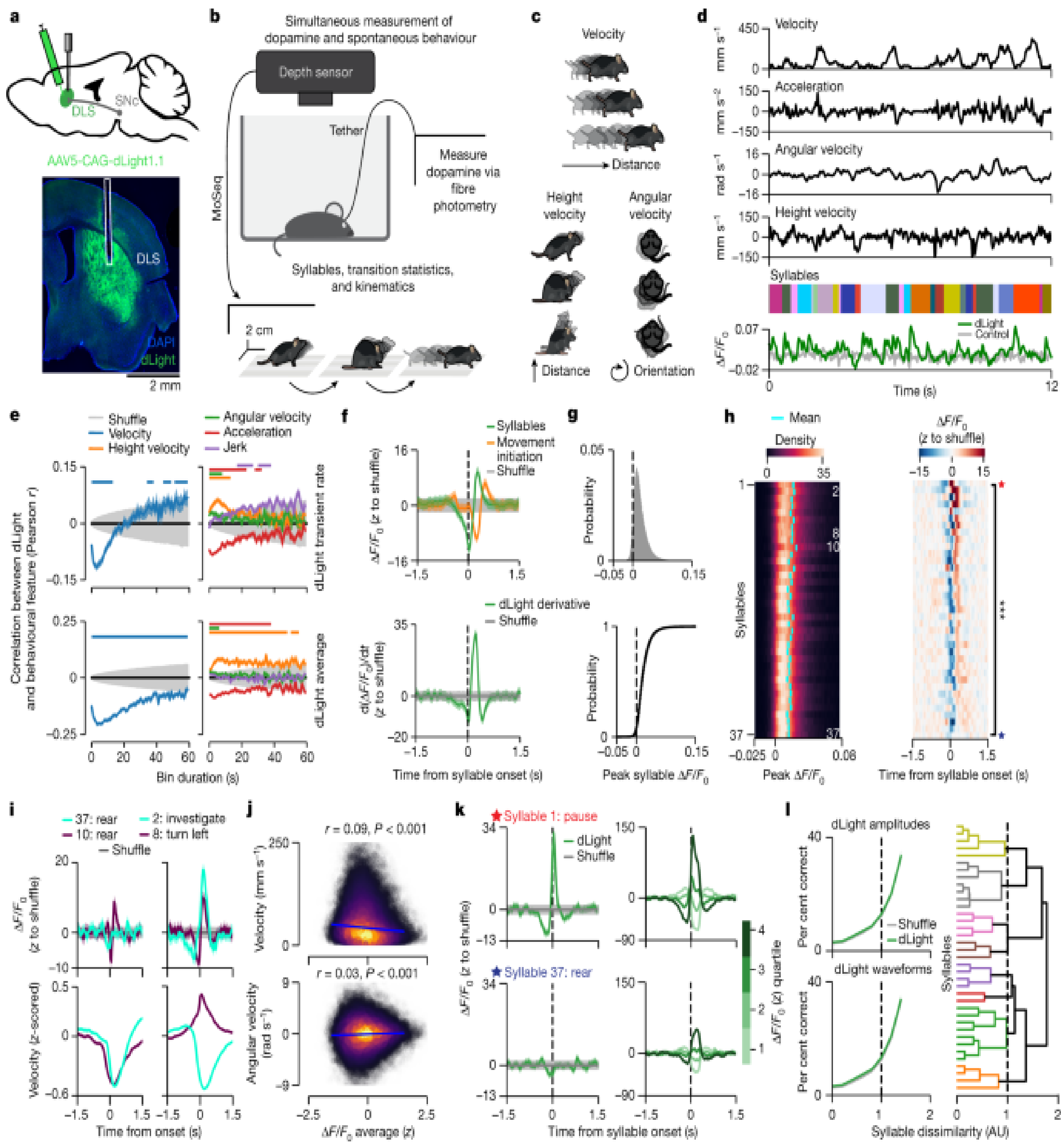
spiny projection neurons (SPNs) and induce plasticity in corticostriatal synapses[30], DLS dopamine fluctuations may be particularly relevant to syllable expression and/or sequencing.

We find that DLS dopamine systematically fluctuates during the expression of behavioural syllables, and—through calibrated closed-loop manipulations of DLS dopamine—demonstrate that these fluctuations are causally related to syllable usage, sequencing and vigour. A simple computational model in which syllable-associated dopamine transients shape sequence composition effectively reproduces the observed behavioural choices made by mice during spontaneous behaviour. These results reveal that DLS dopamine transients act as a continuous teaching signal, one that affords spontaneous behaviour its moment-to-moment structure; our observations further suggest a broad model in which the composition of spontaneous behaviour from elemental components is supported by the same computations and circuits that govern action choices in more structured tasks.

Relating DLS dopamine to spontaneous behavior

To characterize the relationship between striatal dopamine release and spontaneous behaviour, we virally expressed the dopamine reporter dLight1.1 in DLS neurons, and then assessed dopamine fluctuations via photometry as mice explored a featureless open field in the dark[31]. In concert with these neural measurements, we both quantified conventional movement parameters and performed MoSeq[4] (Fig. 1a–d and Extended Data Fig. 1). In this setting, MoSeq identified 37 commonly used behavioural syllables, whose median duration was 400 ± 636 ms, ranging from pause syllables in which mice adopted different static poses, to dynamic syllables in which mice reared, groomed or explored (Extended Data Fig. 2a–c). Syllables varied in terms of how often they were used and the order in which they occurred as time unfolded during each experiment (Extended Data Fig. 2b,d–g).

Fig. 1: Behaviour is associated with dopamine transients in DLS.



a, dLight expression and fibre placement in DLS (Methods). **b**, The behavioural characterization pipeline using MoSeq ($n = 14$ mice for MoSeq, 216 experiments; Methods). **c**, Examples of measured kinematic variables. **d**, Aligned kinematic variables, MoSeq syllables and dLight fluorescence from an example experiment. **e**, Top, average correlation between kinematic variables and dLight transient rate. Bottom, correlations with dLight fluorescence. Coloured shading denotes bootstrapped s.e.m.; grey shading indicates the 95% shuffle confidence interval. Solid bars indicate statistical significance at $P < 0.05$ (shuffle test; Methods). **f**, Top, average fluorescence

(*z*-scored to shuffle; Extended Data Fig. 4c and Methods) aligned to movement initiation or syllable onsets ($n = 100$ shuffles). The average syllable-associated dLight transient exceeds that associated with movement initiation ($P = 0.0006$, $z = 3$, effect size $r = 0.8$, two-sided Wilcoxon signed-rank test; $n = 14$ averages). Bottom, derivative of top panel. Green shading represents the 95% bootstrap confidence interval; grey shading represents the 95% shuffle confidence interval. **g**, The distribution of all syllable-associated dLight peaks. Bottom, the cumulative distribution. **h**, Left, the distribution of syllable-associated dLight peaks for across all experiments. Right, *z*-scored average syllable-associated waveforms, sorted by peak fluorescence. The blue and red stars indicate the syllable waveforms shown in **k**. ***, Kruskal–Wallis *H* test on average syllable-associated fluorescence amplitudes: $P < 10^{-25}$, $H = 209.29$, $n = 518$ mouse–syllable pairs. *y*-axis syllable sorting is shared across panels. **i**, Left, example syllables with different average (across experiments) waveforms (top) but similar velocity (bottom). Right, example syllables with similar waveforms (top) but different velocities (bottom). Shading represents the 95% confidence interval. **j**, Robust linear regression between syllable-associated dLight and velocity (top) or angular velocity (bottom; Methods). Each point is a sampled syllable instance ($n = 28{,}000$ points; $n = 2{,}000$ points per syllable drawn randomly from each mouse). Regression line (shading indicates the 95% bootstrap confidence interval) and kernel density estimate are shown. *P*-values estimated by shuffle test. **k**, Left, average syllable-associated waveforms for starred syllables in **h** (right). Coloured shading represents the 95% bootstrap confidence interval; grey shading represents the 95% shuffle confidence interval. Right, syllable-associated waveforms from the left panel binned into quartiles of peak magnitudes. **l**, Left, held-out classifier performance predicting syllable identity from syllable-associated dLight peak amplitudes (top) or waveforms (bottom). Right, dendrogram showing syllables organized by MoSeq distance (Methods). AU, arbitrary units.

Photometry revealed pervasive, fast dopamine fluctuations in DLS that occurred throughout each experiment regardless of whether the mouse was actively moving or relatively still (average rise time during spontaneous behaviour = 67 ms, decay time = 100 ms; Fig. 1d and Extended Data Fig. 1e,h). These dopamine fluctuations (measured as both dLight transient rates and average binned amplitudes) were only weakly correlated with many

aspects of movement kinematics (for example, turning, rearing and acceleration) but were significantly negatively correlated with translational velocity at short timescales (<10 s); this correlation reversed and became positive at longer timescales, consistent with the idea that dopamine broadly invigorates and motivates movement[6,9,17,32] (Fig. 1e). We validated the relationship between translational velocity and dopamine transients using 3D keypoint tracking, which also demonstrated that forelimb movement per se (that is, independent of translation) only negligibly correlated with dopamine transients (Extended Data Fig. 3).

As has been observed previously, dopamine systematically fluctuated when mice initiated a movement after pausing in the arena[9] (Fig. 1f). However, we also observed that dopamine fluctuated as mice transitioned from one behavioural syllable into the next. These fluctuations exhibited a characteristic dip-then-peak pattern surrounding each syllable transition; as dLight fluorescence changes lag dopamine release by tens of milliseconds[31] (Extended Data Fig. 1e), it is likely that the observed dopamine dip occurs at the end of the previous syllable, and the peak in dopamine occurs during expression of the subsequent syllable (Fig. 1f,g). Consistent with this hypothesis, time-warping the dopamine trace to accommodate differences in syllable duration revealed that on average, dopamine peaked near the middle of each syllable and decayed as the syllable ended, reminiscent of the 'burst–pause' firing pattern of dopamine neurons previously observed at movement initiations[10,11,14,33] (Extended Data Fig. 4a). Nearly every syllable instance was accompanied by a positive dopamine transient, whose amplitude varied (Extended Data Fig. 4b–e).

DLS dopamine does not predict syllable identity

Syllable-specific dopamine waveforms exhibited relatively stereotyped shapes and amplitudes (when averaged either within or across mice and experiments), suggesting that dopamine waveforms predict either the identity of the associated syllable or the kinematics of movement associated with its expression (Fig. 1h and Extended Data Fig. 4c–h). However, kinematically similar syllables (for example, two rears) could exhibit very different average dLight waveforms, whereas kinematically different syllables (for example, turning and investigating) could exhibit similar

average waveforms (Fig. 1i). Aggregating different syllables into categories (such as rearing, grooming or diving) revealed that different categories of behaviour exhibited broadly overlapping average dopamine transient amplitudes (Extended Data Fig. 5a). Furthermore, syllable-associated dopamine transient amplitudes only weakly correlated with the movements actually expressed during a given syllable (Fig 1i,j and Extended Data Fig. 5b).

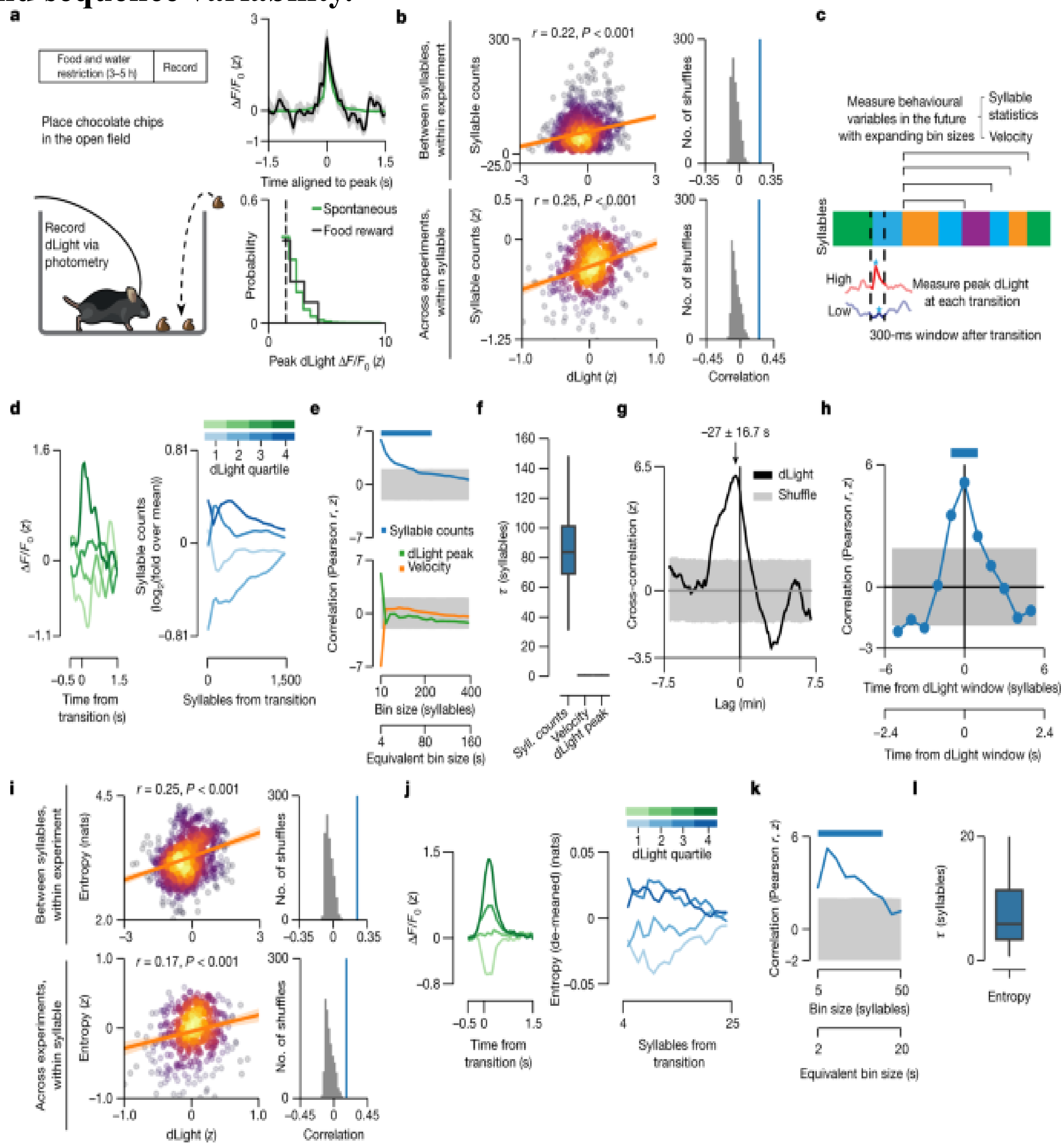
Consistent with a potential dissociation between DLS dopamine and syllable-related kinematics, dopamine waveforms often changed shape and amplitude across different instances of the same syllable (Fig. 1k). This was true even for syllables known to correlate with high SNc and striatal activity such as contralateral turns[20,34,35] (Extended Data Fig. 5c). Indeed, random forest classifiers were unable to predict which syllable was being expressed by the mouse at any moment based on the coincident dLight waveform amplitudes or shapes[9] (Fig. 1l). Syllable-associated dopamine transient amplitudes were also unrelated to many other features of behaviour, including the rendition quality of each syllable, differences in speed between syllables, the biomechanical difficulty of transitioning into a given syllable, the position of the mouse in the arena, and the specific identity of the preceding syllable (Extended Data Fig. 5d–i). Thus, although dopamine systematically fluctuates during the expression of behavioural syllables and syllables are associated on average with different dopamine waveform shapes and amplitudes, individual dopamine transients do not appear to reliably encode information about syllable identity, kinematics or related features of syllable-associated behaviour.

Dopamine predicts future syllable use and sequencing

We therefore considered whether DLS dopamine transients instead have a more flexible role in specifying ongoing patterns of syllable usage, given that the usage of specific syllables and sequences varies during the course of each experiment (Extended Data Fig. 2e–g). Notably, dopamine transient amplitudes observed during spontaneous behaviour roughly matched those observed while mice consumed an unexpected (and presumably rewarding) chocolate chip placed in the open field; this finding suggests that syllable-associated dopamine transients occurring during self-initiated behaviour,

even in the absence of reward, may reinforce the expression of associated syllables (Fig. 2a).

Fig. 2: Endogenous dopamine transients predict average syllable use and sequence variability.



a, The unexpected food reward protocol. Top right, average spontaneous versus food reward-associated transients (Methods). Shading represents bootstrap s.e.m. Bottom right, probability density function of dLight transient amplitudes. The dotted line indicates the threshold for detecting

dLight transient peaks. **b**, Left, robust linear regression between syllable-associated dLight and average syllable counts per syllable (each dot is a syllable–mouse pair). Regression line and kernel density estimate are shown (*r* is Pearson correlation between held-out predictions and actual data). Right, the distribution of Pearson correlations using models fit to shuffled data compared with the observed correlation (blue line). Shading indicates the 95% bootstrap confidence interval. *P*-values estimated by one-sided shuffle test. **c**, Schematic depicting the hypothesis that dopamine predicts changes in future behaviour. Blue star indicates the syllable-associated dopamine peak. **d**, Left, average dLight waveforms for each fluorescence quartile at syllable onset for an example syllable–experiment pair. Right, $\log_2$ fold change compared to average syllable counts after example syllable onset computed over increasing bin sizes (in syllables) after onset. **e**, The average Pearson correlation between syllable-associated dLight and syllable counts or velocity, and the dLight signal autocorrelation, computed using a set of increasing bin sizes after syllable onset. Grey shading represents the shuffled 95% confidence interval. The two *x*-axes reflect time in syllables and approximated in seconds. Solid bars indicate statistical significance ($P < 0.05$, one-sided shuffle test). **f**, The distributions of exponential decay timescales (τ) for the correlations plotted in Fig. 2e ($n = 1{,}000$ bootstrap samples). In all box plots in this Article, the horizontal line represents the median, box edges delineate the first and third quartiles, and whiskers include the furthest data point within 1.5 times the interquartile range of the first or third quartile. **g**, Average cross-correlation between binned syllable counts and syllable-associated dLight fluorescence (from all mice and experiments) across lags ($P < 0.001$, one-sided shuffle test; the arrow indicates average peak lag, error is 68% confidence interval). Grey shading represents the shuffled 95% confidence interval. **h**, Overall correlation between syllable-associated dLight and syllable usage for syllables temporally adjacent to the index syllable. Grey shading represents the shuffled 95% confidence interval. The solid bar denotes statistical significance ($P < 0.05$, one-sided shuffle test). **i**, As in **b**, but for average entropy per syllable. Nat, natural unit of information. **j**, As in **d**, but for sequence entropy for an example syllable–experiment pair. **k**, As in **e**, but for sequence entropy. **l**, Fitted τ values for the correlation curve in **k** ($n = 1{,}000$ bootstrap samples).

Consistent with this hypothesis, those syllables that were on average associated with more DLS dopamine during a given experiment tended to be used more during that experiment; furthermore, variation in the average level of DLS dopamine associated with a given syllable across experiments predicted variation in the use of that same syllable (relative to others) across experiments (Fig. 2b). Dopamine transient amplitudes also predicted the near-term use of associated behavioural syllables: if a given instance of a syllable coincided with a relatively high amplitude dopamine transient, that syllable was then used more over the next several minutes, whereas if an instance of that same syllable coincided with a relatively low amplitude transient it was used less (Fig. 2c–g). The relationship between dopamine transient amplitudes and syllable usage was not due to autocorrelation in the dLight signal or to correlations between dopamine and velocity, both of which declined sharply after a few hundred milliseconds (Fig. 2e). Furthermore, the possible consequences of dopamine transients were largely restricted to the specific syllable with which they were associated, as the size of a given dopamine transient did not substantially correlate with the use of syllables neighbouring in time (Fig. 2h).

Each syllable expressed during spontaneous behaviour is associated with a specific set of possible subsequent syllables whose likelihoods vary substantially—some subsequent syllables are very likely, and therefore contribute to creating predictable (that is, more deterministic) behavioural sequences, whereas others are much less likely and therefore participate in less predictable (that is, more probabilistic) behavioural sequences (Extended Data Fig. 2d). Given that treatment with dopaminergic agonists has been shown to increase the variability of ongoing behavioural sequences[36,37,38], we tested whether syllable-associated dopamine transients can influence the choice of next expressed syllable. Examining the transition patterns between syllables revealed that mice tended to string together more deterministic behavioural sequences when average syllable-associated dopamine was relatively low rather than high (Fig. 2i and Extended Data Fig. 6a). Furthermore, syllable-associated dopamine levels correlated with sequence predictability on a moment-to-moment basis: if a given instance of a syllable was associated with a relatively high amplitude dopamine transient, the mouse tended to make less predictable syllable choices over the next several seconds, whereas if that same syllable was associated with a

relatively low amplitude transient, syllable sequences were more deterministic in the near future (Fig. 2j–l).

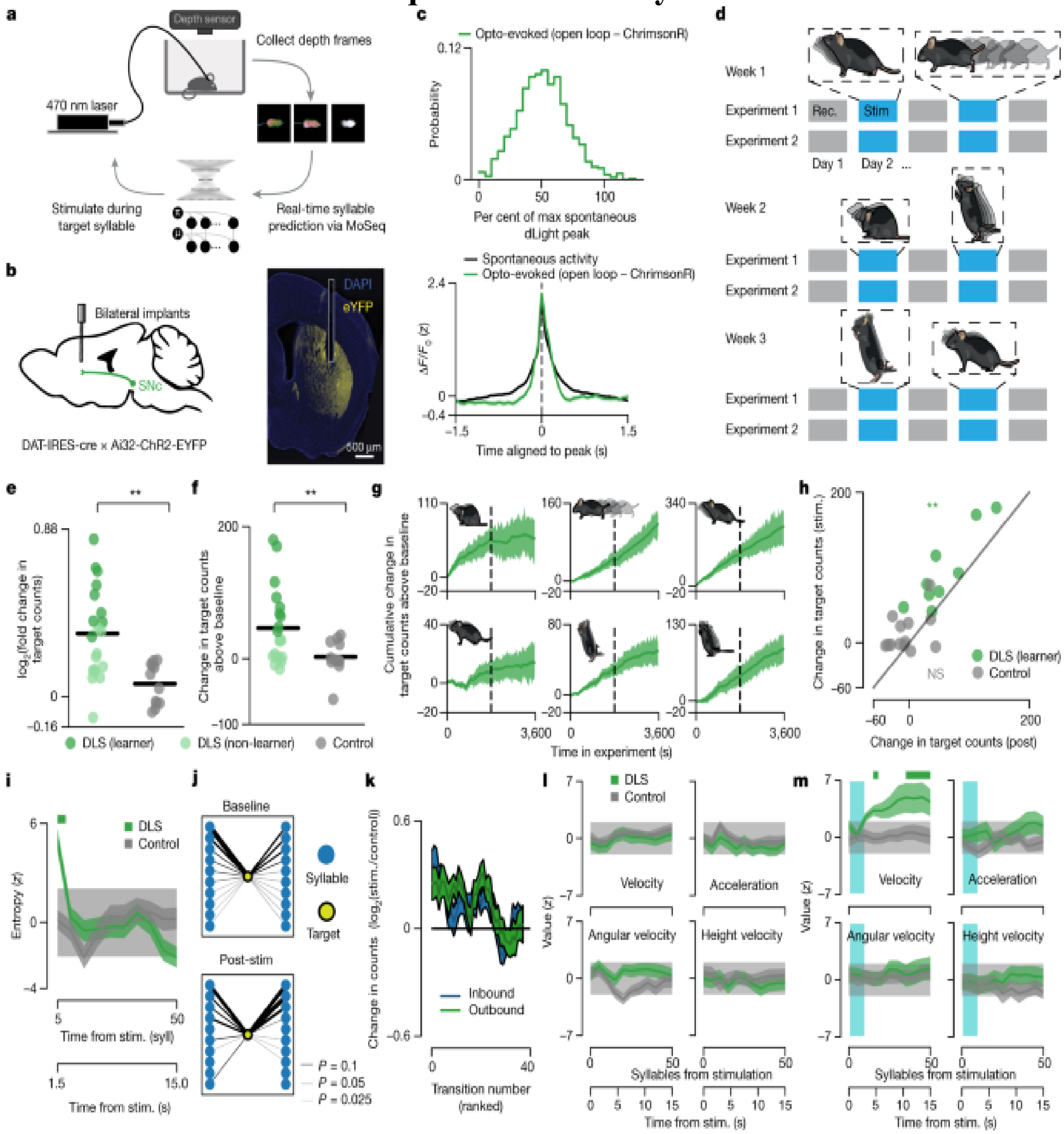
Both syllable usage and sequence variability were themselves correlated: those syllables that were used the most also tended to participate in the most variable behavioural sequences (Extended Data Fig. 6a). However, syllable usage and sequence variability contributed independently to the ability of an encoding model to predict dopamine fluctuations during behaviour (Extended Data Fig. 6b–g; Methods). Our findings regarding syllable usage and sequence variability were specific to DLS, as dLight recordings in dorsomedial striatum (DMS) revealed less frequent dopamine transients that do not predict future syllable usage (Extended Data Fig. 7). However, dLight signals were effectively predicted by velocity in both DMS and DLS, consistent with previous findings that dopamine fluctuations in dorsal striatum correlate with movement[25] (Extended Data Fig. 7e). Taken together, these findings suggest that moment-to-moment fluctuations in DLS dopamine influence the usage of associated behavioural syllables over timescales of minutes, and the choice of what to do next—and thus the predictability of behavioural sequences—on timescales of seconds.

Syllable-associated Opto-DA influences behaviour

To test directly whether syllable-associated DLS dopamine is sufficient to drive increases in syllable usage and sequence variability—and to determine whether fast dopamine fluctuations also influence movements per se—we built a platform that enables us to trigger an optogenetic pulse during the expression of a specific, targeted syllable (Methods and Extended Data Fig. 8a–e). We used this approach to manipulate syllable-associated phasic dopamine levels (Opto-DA) by optically stimulating dopamine axons in the DLS in mice expressing channelrhodopsin-2 (ChR2) in all dopamine-releasing neurons in the midbrain (Fig. 3a,b; Methods). Our stimulation protocol was calibrated to mimic typical syllable-associated dLight amplitudes observed during spontaneous behaviour (Fig. 3c; Methods). We assessed spontaneous mouse behaviour before, during and after syllable-specific stimulation; in each mouse we serially repeated this stimulation protocol for six different target syllables, which collectively exhibited substantial variability in syllable kinematics, usage and sequencing (Fig. 3d

and Extended Data Fig. 9a–e). In nearly all instances, optogenetic stimulation was limited to the targeted syllable and did not extend into subsequent syllables (Extended Data Fig. 9f).

Fig. 3: Optogenetically evoked dopamine release in DLS reinforces syllable use and increases sequence variability.



a, The closed-loop MoSeq pipeline. **b**, Schematic and representative brain section of fibre cannulae over DLS dopamine axons. **c**, Top, normalized optogenetically evoked dLight peak magnitude distribution ($n = 842$ peaks).

Bottom, mean waveforms from spontaneous and Opto-DA transients (Methods). Shading represents the 95% bootstrapped confidence interval. Max, maximum. **d**, Experimental schedule describing baseline (rec.) and stimulation (stim.) sessions for 'target' syllables (Methods). **e**, $\log_2$ fold change in target counts compared with baseline, per mouse, averaged across targets ($P = 0.002$, $U = 197$, $f = 0.82$, one-sided Mann–Whitney U test) (see Methods for definition of 'learner'). **f**, Cumulative increase in target counts relative to baseline in Opto-DA and control mice ($P = 0.007$, $U = 184$, $f = 0.77$, one-sided Mann–Whitney U test). **g**, Cumulative counts over concatenated stimulation sessions per target. Shading indicates bootstrap s.e.m. **h**, The relationship between target syllable usage changes from baseline during stimulation experiments versus post-stimulation experiments per mouse (Pearson $r = 0.89$, $P = 0.005$ for learners and $P = 0.082$ for controls, one-sided shuffle test). NS, not significant. **i**, Average transition entropy following stimulation. Shading indicates bootstrap s.e.m. The light grey band indicates the 95% confidence interval of the pre-stimulation average. Data are binned using five-syllable-wide non-overlapping bins. The bar indicates significance ($P < 0.05$ for Opto-DA mice, $P > 0.05$ for controls, two-sided Mann–Whitney U test comparing stimulation with catch trials). Syll, syllable. **j**, Sequence context changes from baseline to post-stimulation for an example mouse–target pair. Sequences proceed from left (incoming syllables) to right (outgoing syllables). Nodes are sorted by decreasing frequency at baseline. **k**, Average change in inbound and outbound transitions for target syllables on stimulation day sorted by the baseline rank of the transition. Traces are smoothed with a five-point rolling average. Shading indicates bootstrap s.e.m. **l**, Average kinematic parameters aligned to stimulation in Opto-DA mice and controls. Shading as in **i**. No comparisons between stimulation and catch trials in any of the mice were significant ($P > 0.05$, one-sided Mann–Whitney U test). **m**, As in **l**, but following 3-s-long stimulation. The solid bar indicates significance ($P < 0.05$, one-sided Mann–Whitney U test).

Pairing syllable expression with Opto-DA rapidly increased the use of target syllables, evoking a stable increase in syllable usage per unit time (rather than continuously increasing the rate of syllable use) (Fig. 3e–g and Extended Data Fig. 9g,h). This increased usage persisted in experiments after stimulation was terminated, demonstrating that mice learned an

association between dopamine stimulation and the specific targeted syllable (Fig. 3h). The reinforcing effects of Opto-DA were specific to the targeted syllable, as non-target syllables were not affected (Extended Data Figs. 8f and 10). Thus, Opto-DA is sufficient to reinforce the expression of targeted behavioural syllables in the absence of task structure or external sensory cues, suggesting that mice can recognize their own movements on a moment-to-moment basis and use this information to upregulate actions that are associated with exogenous dopamine[39].

Consistent with the observed correlations between endogenous dopamine fluctuations and sequence variability, behavioural sequences observed immediately after Opto-DA stimulation were more unpredictable than the those observed on catch trials (Fig. 3i). Of note, this increase in sequence variability was transient and only apparent during Opto-DA sessions; in the two subsequent sessions—after optogenetic stimulation had ceased but during which target syllable use remained upregulated—behavioural sequences surrounding the target syllable became more predictable, as the most likely transitions into and out of the target became even more likely (Fig. 3j,k). Thus Opto-DA increases sequence variability over seconds-long timescales, whereas sequence variability decreases over the longer timescales at which Opto-DA supports syllable reinforcement.

We also considered whether pairing Opto-DA with specific syllables changed the vigour with syllables were expressed, given prior experiments demonstrating that dopamine can control the speed (that is, vigour) of future movements by reinforcing the expression of fast (or slow) versions of a given movement, or the pitch of a note in a zebra finch's song[40,41,42,43]. To address whether Opto-DA influences syllable-associated vigour, we tailored our Opto-DA experiment such that optogenetic stimulation was delivered only on those target syllable instances in which syllable speed was in the top quartile of its overall distribution; this manipulation systematically increased the velocity with which the target syllable was later expressed (Extended Data Fig. 10d). Conversely, Opto-DA during the slowest quartile of the syllable velocity distribution caused future instances of the targeted syllable to slow down.

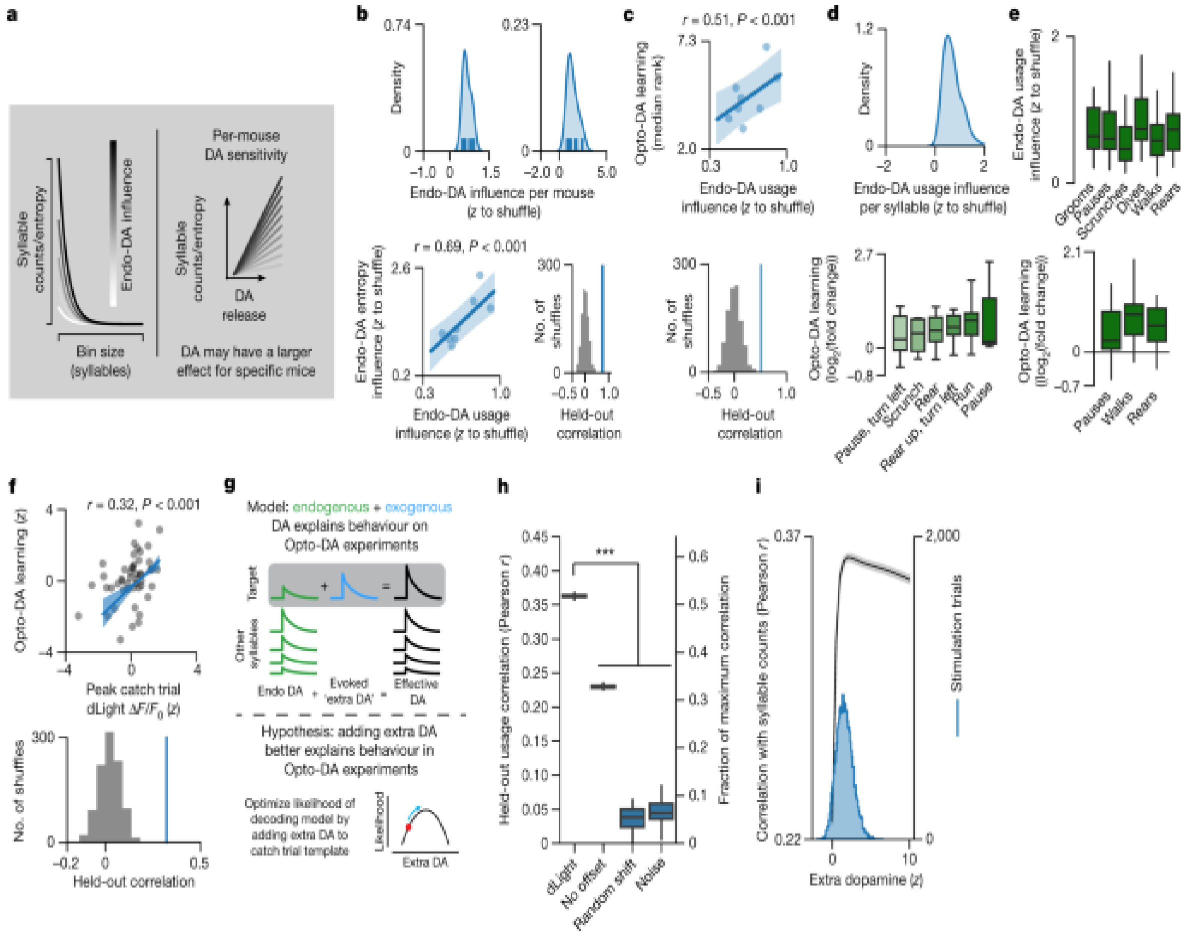
In contrast to its dynamic effects on syllable usage, sequence variability and syllable vigour, Opto-DA did not prompt switching from one syllable to the

next, alter movement parameters associated with the targeted syllable, change movement during the stimulation experiment in general, or induce a preference for a spatial location in the arena (Fig. 3l and Extended Data Fig. 10e–k). However, extending the duration of optogenetic stimulation to several seconds (which caused dLight signals that are of substantially higher amplitude than typically observed during spontaneous behaviour) caused mice to increase their velocity, consistent with previous reports demonstrating that strong perturbations of dopamine are sufficient to cause movements[9,10,14,20] (Fig. 3m and Extended Data Fig. 10l).

Dopamine–behaviour relationships predict learning

Only a fraction of mice successfully associated Opto-DA with targeted syllables, and among those that learned this association, the degree of learning varied (Fig. 3e). We wondered whether this distribution in learning reflected differences in the sensitivity of individual mice to dopamine (Fig. 4a). Consistent with this possibility, the ability of endogenous, syllable-associated dopamine transients to support syllable use and induce sequence variability was similar within an individual mouse but differed across mice (Fig. 4b). Furthermore, mice that were strongly influenced by endogenous dopamine fluctuations were also those that were particularly avid learners of the association between Opto-DA and targeted syllables (Fig. 4c). These findings suggest that inter-mouse variability in Opto-DA learning reflects differences in the sensitivity of DLS to dopamine transients in general (as reflected by behaviour).

Fig. 4: Optogenetic syllable reinforcement varies predictably across mice and syllables.



a, Schematic depicting the relationship between observed endogenous dopamine-syllable usage correlations and per-mouse dopamine sensitivity. Dopamine (DA) sensitivity refers to the ability of endogenous, syllable-associated dopamine peaks to influence changes in syllable counts (Endo-DA count influence) or sequencing (Endo-DA entropy influence) within an experiment (see Methods for how indices were computed). **b**, Top, the distribution of per mouse Endo-DA count influence (left) and Endo-DA entropy influence (right) averaged across all syllables. Bottom, scatter plot (including linear regression model fit) of per mouse average Endo-DA count influence and Endo-DA entropy influence (Pearson $r = 0.69$ computed from model predictions on leave-two-out held-out data; $P = 0.001$, P-value computed via one-sided shuffle test). Shading indicates the 95% bootstrap confidence interval. **c**, Scatter plot (including regression line) of per mouse average Opto-DA learning versus Endo-DA count influence (Pearson $r =$ 0.51 (computed from model predictions on leave-two-out held-out data); P = 0.001, P-value computed via one-sided shuffle test). Shading indicates the

95% bootstrap confidence interval. **d**, Top, the distribution of per syllable average Endo-DA count influence ($n = 296$ mouse–syllable pairs). Bottom, Opto-DA learning plotted syllable-by-syllable for 'learner' mice ($n = 9$ mice). **e**, Top, average Endo-DA count influence across syllable categories. Bottom, average Opto-DA learning across syllable categories. **f**, Top, scatter plot (including regression line) of catch trial syllable-associated dLight and Opto-DA learning for each mouse–syllable pair ($r = 0.32$ over held-out data; $P < 0.001$, estimated via one-sided shuffle test). Bottom, model performance (evaluated with five times fivefold cross-validation) using actual versus shuffled data. **g**, Hypothesis that evoked dopamine release combines with ongoing endogenous release to alter behavioural choices. **h**, Model-based likelihood of predicting held-out syllable choices on Opto-DA stim experiments (blue) relative to control models (Methods) ($P = 7 \times 10^{-18}$ across all model comparisons relative to dLight model, $U = 2{,}500$, $f = 1$, two-sided Mann–Whitney U test; $n = 50$ model restarts; Methods). The right y-axis indicates the model performance as a fraction the of maximum correlation. **i**, The relationship between average model accuracy (correlation between predicted syllable usage and actual usage) and the 'extra dopamine' free parameter (black). Shading indicates the 95% bootstrap confidence interval. The distribution of empirically measured optically evoked dLight fluorescence is shown in blue.

In addition, not all syllables were equally reinforceable in both the endogenous dopamine and Opto-DA experiments (Fig. 4d). There was no discernable relationship between the specific type of behaviour encoded by each syllable (for example, rearing, running or grooming) and the degree of reinforcement by endogenous dopamine or Opto-DA (Fig. 4e). By contrast, the ability of a particular syllable to be reinforced by exogenous dopamine was well predicted by that syllable's average endogenous dopamine transient amplitude: syllables that on average were associated with high amplitude dopamine transients during spontaneous behaviour were more easily reinforced by Opto-DA than those typically associated with low amplitude dopamine transients (Fig. 4f).

This observation suggests that the endogenous dopamine fluctuations that naturally occur during expression of a targeted syllable sum together with the exogenous dopamine induced by Opto-DA to promote syllable usage. To test this idea, we built a dynamic decoding model that predicted moment-to-

moment syllable expression on the basis of syllable-associated dopamine. In this model, syllable-associated dopamine was represented as the sum of two components: endogenous dopamine (that is, the observed syllable-associated dopamine observed at baseline) and a free parameter representing the extra dopamine afforded by Opto-DA. We then fit this model to data observed during catch trials in the Opto-DA experiments, in which optogenetic stimulation during the target syllable was not provided (Methods). This additive model could reliably predict moment-to-moment syllable choices during the Opto-DA experiment (which included stimulation trials), but made less effective predictions when dopamine transients were shuffled in time relative to syllables (Fig. 4g,h). In addition, the amount of extra dopamine required for the model to accurately predict actual syllable usage patterns closely matched the amount of dopamine elicited by Opto-DA in our calibration experiments (Figs. 3c and 4i). Together, these data support a model in which endogenous and exogenous DLS dopamine sum together to influence syllable expression, and provide a possible explanation for the differential reinforcement of syllables observed in the Opto-DA experiment.

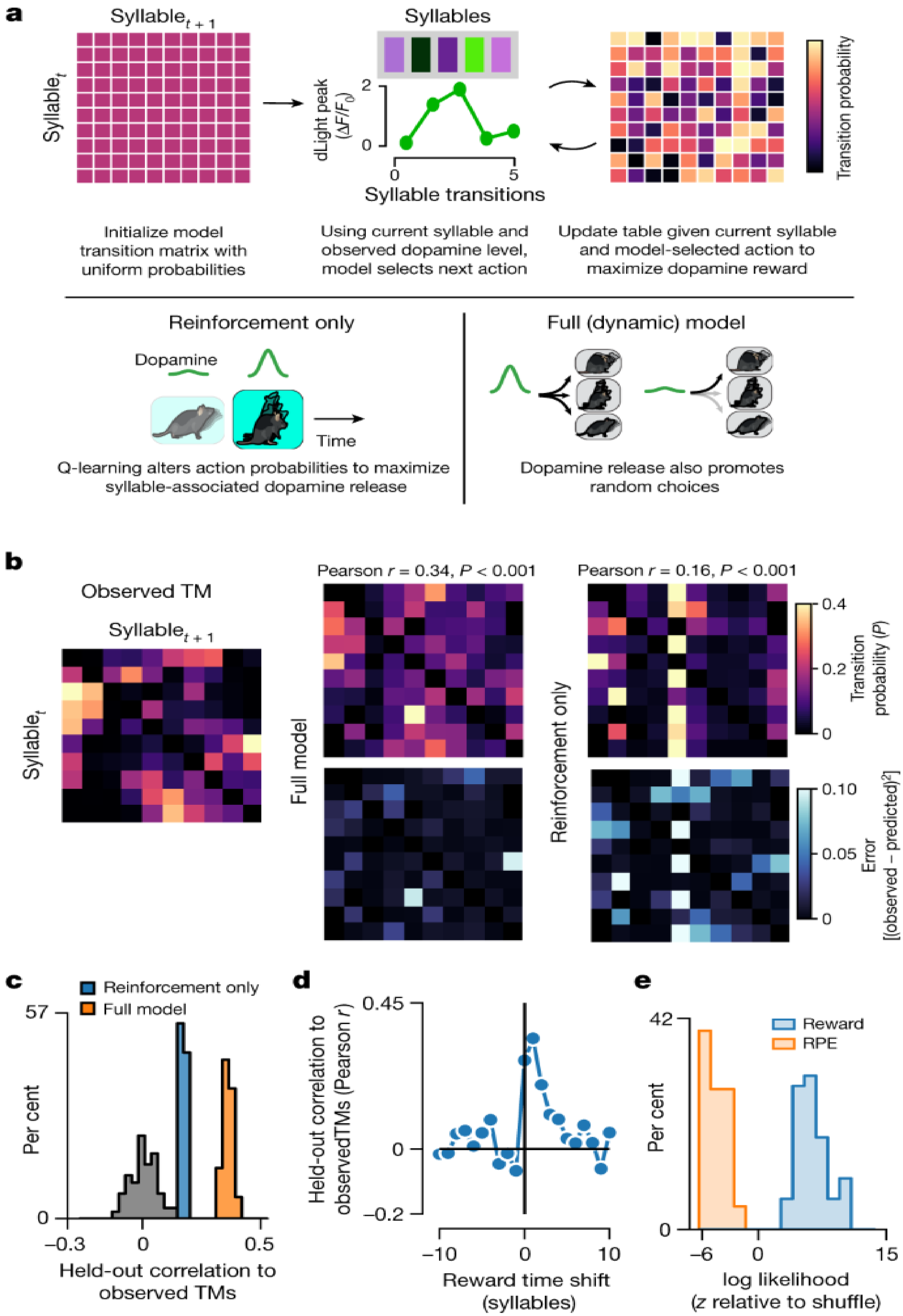
An RL model reproduces spontaneous behaviour

The observation that both exogenously induced and endogenous fluctuations in DLS dopamine levels drive changes in syllable usage and sequencing raises the surprising possibility that, during spontaneous exploration, mice are optimizing their behaviour—as they do in more structured tasks—to maximize the amount of total dopamine obtained during an experiment. If so, this would suggest that mice interpret fast DLS dopamine fluctuations during spontaneous behaviour as a teaching signal capable of shaping action choices.

To test this hypothesis, we built a reinforcement learning (RL) model in which an agent is trained—based on observed syllable sequences and dopamine transients—to predict the syllable choices expressed by actual mice during spontaneous behaviour (Fig. 5a). Whereas RL agents typically seek to maximize overall reward by optimizing their action choices in a particular state, here the RL agent seeks to maximize dopamine by optimizing the choice of which syllable to express next given its current syllable (Methods). We train this RL agent using standard Q-learning rules,

which govern how syllable-associated dopamine transients update the transition probabilities between syllables[44], and formulated this model such that dopamine is both rewarding and injects transient variability into subsequent syllable choices.

Fig. 5: RL models suggest that mice attempt to maximize dopamine during spontaneous behaviour.



a
Syllable$_{t+1}$
Syllable$_t$
Syllables
dLight peak ($\Delta F/F_0$)
Syllable transitions
Transition probability
Initialize model transition matrix with uniform probabilities
Using current syllable and observed dopamine level, model selects next action
Update table given current syllable and model-selected action to maximize dopamine reward
Reinforcement only
Dopamine
Time
Q-learning alters action probabilities to maximize syllable-associated dopamine release
Full (dynamic) model
Dopamine release also promotes random choices
b
Observed TM
Syllable$_{t+1}$
Syllable$_t$
Pearson r = 0.34, P < 0.001
Pearson r = 0.16, P < 0.001
Full model
Reinforcement only
Transition probability (P)
Error [(observed – predicted)^2]
c
Reinforcement only
Full model
Per cent
Held-out correlation to observed TMs
d
Held-out correlation to observedTMs (Pearson r)
Reward time shift (syllables)
e
Reward
RPE
Per cent
log likelihood (z relative to shuffle)

a, Top, schematic describing modification of a standard RL model to explore relationships between DLS dopamine fluctuations and behavioural choices. Bottom, schematics of ‘reinforcement-only’ and ‘full’ model variants (Methods). **b**, Left, empirical transition matrix (TM) observed during open field behaviour. Centre, an example transition matrix learned by the full model (top), along with the squared difference between the empirically observed transition matrix and the example learned transition matrix (bottom). Right, as in centre, except for the reinforcement-only model. The average correlation between the observed transition matrix and the transition matrix learned by each model, along with the associated *P*-value computed via shuffle test, are given for each model type. For visualization, the model transition matrix is estimated by taking a softmax (see Methods) over the Q-table learned by the model. Here, the temperature parameter was set to 0.1 for visualization only. **c**, The distribution of correlations between the learned and observed transition matrices for both the reinforcement-only (blue) and full (orange) models, compared to a histogram of correlations between transition matrices learned with time-shuffled dLight traces (all models are statistically significant according to a shuffle test, defined as model fits exceeding 95% of shuffle correlation values, $n = 100$ shuffles). **d**, The performance of the full model after temporally shifting syllable-associated dLight amplitudes across syllables over various lags. **e**, The distribution of log likelihoods for models that consider dopamine as a reward versus a reward-prediction error (RPE) signal (Methods). The log likelihoods shown are for the best parameterization for each model type across 50 bootstraps of the dataset. On the basis of this relationship, we formulated models that treated dopamine transients as representing reward rather than reward-prediction error.

Inspection revealed that after training the model converged on a syllable transition matrix similar to that emitted by actual mice exploring an open field; comparing alternative formulations of this model revealed that maximal model performance depended on dopamine both reinforcing specific syllable transitions and briefly increasing the variability of syllable choices (Fig. 5b–d). These findings are consistent with a model in which mice structure spontaneous behaviour to maximize DLS dopamine. We note that although our models assume that dopamine acts as a reward, there are alternative formulations under which dopamine acts as a reward-prediction error that are also consistent with our data (Fig. 5e; Methods). This caveat is

important given that mice rapidly modify syllable usage in response to Opto-DA stimulation, preventing us from formally distinguishing between these possibilities.

Discussion

Goal-oriented behaviours are purposive and yield explicit rewards, whereas spontaneous behaviour can often appear to be aimless and inscrutable. And yet, even the behaviour of mice placed in a dark empty bucket exhibits remarkable structure. Here we show that this structure is governed by ongoing fast fluctuations in the neuromodulator dopamine. DLS dopamine transients both correlate with and causally influence how often each syllable is used and in what order, despite the absence of an explicit task or exogenous reward. The ability of a simple RL model, in which dopamine transients are substituted for the reward signal, to predict syllable choices suggests that dopamine acts as a moment-to-moment teaching signal, one that enables DLS to dynamically assemble behavioural sequences. Together, our observations identify a neural mechanism that actively shapes the trajectory of spontaneous behaviour as it unfolds, and propose an unexpected functional role for DLS dopamine during self-motivated action that is centred around choice and reinforcement rather than movement initiation or instantaneous kinematic control.

Recordings of dopamine neurons during spontaneous behaviour have revealed a variety of correlations with movement initiation, speed and kinematics, and strong optogenetic stimulation of dopamine neurons elicits movement, suggesting that dopamine may cause movements per se[9,11,12,14,20,33]. The failure of our calibrated optogenetic experiments to causally influence movement initiation or execution at the moment of stimulation argues that DLS dopamine organizes—rather than commands—movement by influencing the overall statistical structure of behaviour. Notably, DLS dopamine bidirectionally influences future syllable vigour, demonstrating that dopamine regulates both discrete (that is, which syllable to express next) and continuous (that is, how fast to express a particular syllable) aspects of spontaneous behaviour.

Our data suggest four broad, non-mutually exclusive models for how DLS dopamine transients may arise and thereby influence future behavioural choices. First, dopamine may represent an output of the motor system that enables it to modulate (and invigorate) the future expression of some syllables relative to others. In this model, DLS dopamine acts to implement a motor plan articulated by the cortex, thalamus and basal ganglia (all of which send projections to SNc[45]); our observations that syllable-associated dopamine causally influences both future syllable use and subsequent syllable choices is consistent with this model. Although the mechanisms that enable dopamine to briefly increase sequence variability are not yet clear, it is possible that this phenomenon relates to the ability of dopamine to increase SPN excitability, which may decrease the fidelity with which cortical inputs are transformed into ensemble SPN activity[30,46].

Second, DLS dopamine may represent the output of a circuit that evaluates the content of spontaneous behaviour. Although dopamine has classically been thought to report reward-prediction errors—which by definition require the provision of reward—it has recently been argued that dopamine may also encode action-prediction errors[47,48] (APEs). APEs are proposed to occur as animals either execute or plan to execute a behaviour that is unexpected in a given context; in the setting of spontaneous behaviour, an APE-like model would predict that DLS dopamine represents the comparison between the expressed (or soon-to-be-expressed) behavioural syllable and that which would have been expressed at a particular moment given an idealized transition matrix. Our finding (similar to that in ref. [9]) that syllable-associated dopamine transients reflect the probability of the next expressed behavioural syllable—but convey no information about syllable identity—is also consistent with the proposed role of APEs in conveying information about action errors that is independent of the specific identity of the expressed behaviour.

Third, DLS DA may encode an error signal unrelated to APEs. Given that most syllables are probably associated with some degree of active sensory sampling, it is possible that DLS dopamine reflects the difference between expected and experienced sensory cues in the environment that are encountered during each syllable. Similarly, DLS dopamine might evaluate the actual quality of execution of each syllable (akin to performance

prediction errors observed in birdsong[49,50]), although our data suggest that this is not the case. Finally, the brain may be misinterpreting random fluctuations in DLS dopamine as a reward-like signal, thus inadvertently structuring behaviour; such stochastic fluctuations in cortical dopamine have recently been shown to support RL[51]. However, spontaneous behaviour in the open field evolves during our experiments with characteristic dynamics, arguing against this possibility. Future work will be required to arbitrate among these models.

We note that the midbrain dopamine system targets many brain areas with a diverse array of functions, and our experiments were deliberately designed to focus on the influence of dopamine on the DLS[14]. Given differences in their anatomical inputs and intrinsic timescales at which dopamine fluctuates[52], it is likely that the relationship that we observe between DLS dopamine and syllable usage does not apply to the DMS or the ventral striatum, which vary in their functional roles in movement, motivation and value assignment[24,53,54]. Indeed, it is possible that phasic dopamine fluctuations in these areas are responsible for initiating new movements or controlling their kinematics; alternatively, it is possible that tonic dopamine is broadly permissive but not causal for movements per se, as is suggested by the ability of l-DOPA therapy to revert movement deficits in humans with Parkinson's disease[55]. The causal relationship between dopamine and movement may also depend on the specific task in which an animal is engaged and the extent of its training; the observation that during goal-oriented tasks many individual dopamine neurons adopt task-related tuning to movements is consistent with this idea[11,16,19]. Note also that although both the endogenous dopamine and calibrated Opto-DA experiments argue that DLS dopamine is unrelated to movement initiation or kinematics (similar to observations in refs. [9,10,11,14,19]), it is possible that our bulk measurements and manipulations obscure a subtle role for specific SNc dopaminergic axons in controlling instantaneous movement.

Despite training, animals often fail to learn to perform structured tasks, with some animals, tasks or behaviours being more resistant to learning than others; understanding brain–behaviour relationships under these circumstances often requires considering only those animals whose behavioural responses exceeds some threshold for accuracy after training[56].

Although our experiments contain no overt goal or task structure, we also observe variability in Opto-DA learning, both across individual mice and across syllables. A substantial part of this variability can be explained by dopamine itself: mice that are more sensitive to endogenous dopamine fluctuations are better able to associate Opto-DA with targeted syllables, and syllables more effectively associated with Opto-DA are also associated with higher average endogenous dopamine transients. Our results also suggest that some mice are more sensitive to dopamine than others, despite being genetically identical and housed in similar conditions; further work is required to characterize the source of this inter-mouse variation.

Previous work has demonstrated that DLS SPNs encode information about the current syllable and the sequence context in which that syllable is expressed[27,28,29]. This information is probably inherited from thalamic and cortical inputs to DLS[57,58]. The ability of mice to upregulate syllables in response to DLS dopamine axon stimulation demonstrates that mice can recognize and reinforce the movements of their own body in the absence of external sensory cues (such as the click commonly used as a sensory trigger during clicker training); this self-recognition is remarkably specific, as dopamine stimulation does not substantially reinforce syllables that are kinematically related to the target, or syllables that are adjacent in time. Our findings naturally lead to a hypothesis in which dopamine encourages the use of associated syllables by inducing short-term plasticity in sensorimotor inputs to DLS[59]. Notably, our dopamine stimulation coincides with only a fraction of each targeted syllable, yet syllables remain temporally intact and are coherently reinforced as a whole; this observation supports the speculation that syllables are natural units of spontaneous behaviour used by the brain to structure action.

Methods

A list of reagents and resources is provided in Extended Data Table 1.

Ethical compliance

All experimental procedures were approved by the Harvard Medical School Institutional Animal Care and Use Committee (Protocol Number 04930) and

were performed in compliance with the ethical regulations of Harvard University as well as the Guide for Animal Care and Use of Laboratory Animals.

MoSeq

Overview

MoSeq (described previously in refs. [4,27,60]) is an unsupervised machine learning method that identifies brief, re-used behavioural motifs that mice perform spontaneously. MoSeq takes as its input 3D imaging data of mice and returns a set of behavioural ‘syllables’ that characterizes the expressed behaviour of those mice, and the statistics that govern the order in which those syllables were expressed in the experiment. MoSeq was used as it is originally described to explore relationships between endogenous DLS dopamine release and behaviour. This technology was further adapted to accommodate real-time syllable identification for closed-loop manipulations of neural activity as described below. Importantly, the underlying fitted autoregressive hidden Markov model (AR-HMM) for both the ‘offline’ and ‘online’ variants of MoSeq used in this study are the same, enabling comparisons of neural activity associated with syllables that were recognized and performed across multiple experiments.

Pre-processing

MoSeq consists of two essential workflows: one for pre-processing depth data and converting it into a low-dimensional time series that describes pose dynamics, and another for modelling the low-dimensional time-series data. As previously described, in order to focus on pose dynamics, raw depth frames were first background-subtracted to convert depth units from distance to height from the floor (in millimeters). Next, the location of the mouse was identified by finding the centroid of the contour with the largest area using the OpenCV findcontours function. An 80 × 80 pixel bounding box was drawn around the identified centroid, and the orientation was estimated using an ellipse fit (with a previously described correction for ±180-degree ambiguities[4,27]). The mouse was rotated in the bounding box to

face the right side. The 80 × 80 pixel depth video of the centred, oriented mouse was then used to estimate pose dynamics.

Size-normalizing deep network

To accommodate noise in online syllable estimation and other sources of variation in the depth images not due to changes in pose dynamics (for example, occluding objects such as fibre-optic cables), we designed a denoising convolution autoencoder. The network was designed using TensorFlow to process images in <33 ms, the time between frame captures on the Microsoft Kinect V2[61]. On the encoder side, 4 layers of 2D convolutions (ReLu activation) followed by max pooling were used to downsample the 80 × 80 images to 5 × 5. Another 4 layers of 2D convolutions with successive upsampling layers were used on the decoding side to reconstruct the 80 × 80 images (10,310,041 total parameters). Batch normalization was used during training with a batch size of 128. In order to train the network, we used a size- and age-matched dataset (7–8 weeks of age). Mouse images were corrupted through rotation, position jitter, zooming in and out (that is, changing size), and superimposing depth images of fibre-optic cables. The network was fed corrupted mouse images as input and was trained to minimize the reconstruction loss of the original, corresponding uncorrupted mouse images (Extended Data Fig. 8a–c). The model was trained for 100 epochs using stochastic gradient descent with early stopping. Both online and offline variants of MoSeq included the size-normalizing network to ensure results were comparable.

Dimensionality reduction and AR-HMM training

In order to represent pose dynamics in a common space for all experiments, principal components and an AR-HMM time-series model were trained offline on a sample dataset of genotype- and age-matched mice. The parameters describing the principal components and AR-HMM model were saved. All depth videos acquired for this paper were then projected onto these same principal components for all experiments, whether they used the online or offline variant. As previously described, principal components were estimated from cropped, oriented depth videos, and the AR-HMM was trained on the top 10 principal components. Since the denoising autoencoder

was used for all experiments, mouse videos from the size-and-age-matched dataset were fed through the denoising autoencoder prior to principal component estimation.

Offline variant

In the offline variant, the Viterbi algorithm was used to estimate the most probable discrete latent state sequence according to the trained AR-HMM for each experiment post hoc. This variant was used to analyse all data except for the Opto-DA experiments shown in Figs. 3 and 4.

Online variant

In the online variant, syllable likelihoods were computed and updated by computing the forward probabilities of the discrete latent states for each frame as they arrived from the depth sensor. To avoid spurious syllable detections, the targeted syllable probability had to cross a user-defined threshold for three consecutive frames.

Histological verification

Mice were euthanized following completion of behavioural tests. Mice were first perfused with cold 1× PBS and subsequently with 4% paraformaldehyde. Fifty-micrometre sections of extracted brain tissue were sliced on a Leica VT1000 vibratome. All slices were mounted on glass slides using Vectashield with DAPI (Vector Laboratories) and imaged with an Olympus VS120 Virtual Slide Microscope.

dLight validation and variant selection

dLight1.1 was selected to visualize dopamine release dynamics in the DLS owing to its rapid rise and decay times, comparatively lower dopamine affinity (so as to not saturate binding), as well as its responsiveness over much of the physiological range of known DA concentrations in freely moving rodents[31,62,63,64].

Since dopamine-free and dopamine-bound excitation spectra have yet to be reported for the dLight1.1 sensor, a series of in vitro experiments was performed to identify an excitation wavelength whose fluorescence was stable and independent of dopamine levels, and which therefore could be used for post hoc motion artefact correction. Like GCaMP, dLight1.1 uses cpGFP as a chromophore, and various generations of GCaMP have been shown to: (1) have an increase in ligand-free fluorescence when excited with 400 nm wavelengths and (2) have an isosbestic wavelength in the UV to blue region[65,66,67]. To test whether UV excitation could be a suitable reference wavelength for dLight1.1, HEK 293 cells (ATCC, cells were validated by ATCC via short tandem repeat analysis and were not tested for mycoplasma) were transfected with the dLight1.1 plasmid (Addgene 111067-AAV5) using Mirus TransIT-LT1 (MIR 2304). Cells were imaged using an Olympus BX51W I upright microscope and a LUMPlanFl/IR 60×/0.90W objective. Excitation light was delivered by an AURA light engine (Lumencor) at 400 and 480 nm with 50 ms exposure time. Emission light was split with an FF395/495/610-Di01 dichroic mirror and bandpass filtered with an FF01-425/527/685 filter (all filter optics from Semrock). Images were collected with a CCD camera (IMAGO-QE, Thermo Fisher Scientific), at a rate of one frame every two seconds, alternating the excitation wavelengths in each frame. Image acquisition and analysis were performed using custom-built software written in MATLAB[68] (Mathworks). Cells were segmented from maximum-projection fluorescence images using Cellpose[69]. Cells with a diameter of less than 30 pixels were excluded from downstream analysis. Fluorescence traces were denoised using a hampel filter (window size 10 and threshold set to 2 median absolute deviations from the median) and normalized to $\Delta F/F_0$. Cells were included if their maximum $\Delta F/F_0$ exceeded 5%. F_0 was computed by fitting a bi-exponential function to the time series.

Stereotaxic surgery for open field photometric recordings

Eight- to ten-week-old C57BL/6J ($n = 6$ mice, The Jackson Laboratory stock no. 000664) mice of either sex were anaesthetized using 1–2% isofluorane in oxygen, at a flow rate of 1 l min^{-1} for the duration of the procedure. AAV5.CAG.dLight1.1 (Addgene #111067, titre: 4.85×10^{12}) was injected at

a 1:2 dilution (either sterile PBS or sterile Ringer's solution) into the DLS (AP 0.260; ML 2.550; DV −2.40), in a total volume of 400 nl per injection. For all stereotaxic implants, AP and ML were zeroed relative to bregma, DV was zeroed relative to the pial surface, and coordinates are in units of mm. Injections were performed by a Nanoject II or a Nanoject III (Drummond) at a rate of 10 nl per 10 s, unilaterally in each mouse. A single 200-µm diameter, 0.37–0.57 NA fibre cannula was implanted 200 µm above the injection site at the DLS (DV −2.20) for photometry data collection. Finally, medical-grade titanium headbars (South Shore Manufacturing) were secured to the skull with cyanoacrylate glue (Loctite 454).

Mice were group-housed prior to stereotaxic surgery procedures, and following surgery were individually housed on a 12-hour dark–light cycle (09:00–21:00). All behavioural recordings were done between 010:00 and 17:00.

Stereotaxic surgery for simultaneous photometric recordings and optogenetic stimulation

Six- to 12-week old DAT-IRES-cre mice (n = 10 mice, The Jackson Laboratory stock no. 006660) of either sex were injected with the same dLight1.1 virus described above into the right hemisphere DLS. Additionally, using the same previously described surgical procedure, 350 nl of AAV1.Syn.Flex.ChrimsonR.tdTomato (UNC Vector Core, titre: 4.1×10^{12}) was injected into the right hemisphere SNc (AP −3.160; ML 1.400; DV −4.200 from pia), in a 1:2 dilution for calibration and stimulation experiments (see below). Mice were implanted unilaterally with a 200 µm core 0.37–0.57 NA fibre over the DLS for simultaneous stimulation and photometric data collection.

Two of the ten mice were used to calibrate optogenetic stimulation (see 'dLight calibration experiments'). The other 8 mice injected with dLight and ChrimsonR were also run through the 3 complete closed-loop experiments described in 'Closed-loop DLS dopamine stimulation experiments' (one experiment with 250 ms continuous wave (CW) stimulation, one with 2 s CW stimulation, and another with 3 pulsed stimulation, 25 Hz frequency with 5 ms pulse width). Baseline data from these experiments were

combined with mice described in ‘Fibre Photometry for dLight recordings’, thus yielding a total of $n = 14$ mice. Two of the 12 dLight only mice did not pass our quality control criteria for dLight recordings and were thus excluded from all dLight analysis (note that they were included in Extended Data Fig. 2a–b,d only, which strictly used behavioural data). Baseline data were considered data from the day prior to a stimulation day, or the day after with the targeted syllable excluded (yielding $n = 378$ experiments total). If the targeted syllable could not be reasonably excluded then data from the day after a stimulation day was excluded entirely.

dLight behaviour procedures

OFA experiments

Depth videos of mouse behaviour were acquired at 30 Hz using a Kinect 2 for Windows (Microsoft) using a custom user interface written in Python (similar to ref. [60]) on a Linux computer. For all OFA experiments, except where noted, mice were placed in a circular open field (US Plastics 14317) in the dark for 30 min per experiment, for 2 experiments per day. As described previously, the open field was sanded and painted black with spray paint (Acryli-Quik Ultra Flat Black; 132496) to eliminate reflective artefacts in the depth video.

Food reward experiments

To assess whether spontaneous dLight transients in the DLS were of appreciable magnitude compared to reward consumption-related transients, a series of separate dLight photometry experiments were run to measure reward consumption-related transient magnitudes ($n = 6$ mice). For two days prior to the experiment, mice were habituated to the open field arena for two 30-min experiments on each day. On the morning of the experiment, to increase the salience of food reward, mice were habituated to the experimental room and food and water restricted for 3–5 h prior to beginning the experiment. Mice were placed in the arena, and behaviour and photometry data were simultaneously acquired. Chocolate chips (Nestle Toll House Milk Chocolate) were divided into quarters and introduced into the arena at random intervals and locations decided by the experimenter (with

an average of 1 chocolate chip piece every 4 min) for mice to freely consume for a total of 30 min. To identify reward consumption-related responses, a human observer indicated each moment in time during the experiment where mice began to consume the chocolate via post hoc inspection of the infrared video captured by the Kinect. Photometry signal peaks for Fig. 2a were identified at the onset of consumption. Mean spontaneous transient peak had observed magnitudes of 2.12 ± 0.80 $\Delta F/F_0$ (z) (n = 5,247 transients). By comparison, mean reward consumption-associated transients had an approximate magnitude of 2.36 ± 0.92 $\Delta F/F_0$ (z) (n = 10 transients).

Fibre photometry for dLight recordings

Photometry and behavioural data were collected simultaneously. A digital lock-in amplifier was implemented using a TDT RX8 digital signal processor as previously described[27]. A 470 nm (blue) LED and a 405nM (UV) LED (Mightex) were sinusoidally modulated at 161 Hz and 381 Hz, respectively (these frequencies were chosen to avoid harmonic cross-talk). Modulated excitation light was passed through a three-colour fluorescence mini-cube (Doric Lenses FMC7_E1(400-410)_F1(420-450)_E2(460-490)_F2(500-540)_E3(550-575)_F3(600-680)_S), then through a pigtailed rotary joint (Doric Lenses B300-0089, FRJ_1x1_PT_200/220/LWMJ-0.37_1.0m_FCM_0.08m_FCM) and finally into a low-autofluorescence fibre-optic patch cord (Doric Lenses MFP_200/230/900-0.37_0.75m_FCM-MF1.25_LAF or MFP_200/230/900-0.57_0.75m_FCM-MF1.25_LAF) connected to the optical implant in the freely moving mouse. Emission light was collected through the same patch cord, then passed back through the mini-cube. Light on the F2 port was bandpass filtered for green emission (500–540 nm) and sent to a silicon photomultiplier with an integrated transimpedance amplifier (SensL MiniSM-30035-X08). Voltages from the SensL unit were collected through the TDT Active X interface using 24-bit analogue-to-digital convertors at >6 kHz, and voltage signals driving the UV and blue LEDs were also stored for offline analysis.

The output of the PMT was then demodulated into the components generated by the blue and UV LEDs. The voltage signal was multiplied by the two driving signals—corresponding to the green emission due separately

to blue and UV LED excitation—and low-passed using a third order elliptic filter (max ripple: 0.1; stop attenuation: 40 dB; corner frequency: 8 Hz). The UV component was used a reference signal.

Synchronizing depth video and photometry

To align photometry and behavioural data, a custom IR led-based synchronization system was implemented. Two sets of 3 IR (850 nm) LEDs (Mouser part # 720-SFH4550) were attached to the walls of the recording bucket and directed towards the Kinect depth sensor. The signal used to power the LEDs was digitally copied to the TDT. An Arduino was used to generate a sequence of pulses for each LED set. One LED set transitioned between on and off states every 2 s while the other transitioned into an on state randomly every 2–5 s and remained in the on state for 1 s. The sequences of on and off states of each LED set were detected in the photometry data acquired with the TDT and IR videos captured by the Kinect. The timestamps of the sequences were aligned across each recording modality and photometry recordings were down sampled to 30 Hz to match the depth video sampling rate. This same mechanism was used to align photometry data to keypoints in Extended Data Fig. 3.

Photometry pre-processing

Demodulated photometry traces were normalized by first computing $\Delta F/F_0$. F_0 was estimated by calculating the 10th percentile of the photometry amplitude using a 5-s sliding window to account for slow, correlated fluorescence changes between dLight and the UV reference channels. Both the dLight and reference channels were normalized using this procedure. Since the UV reference signal captures non-ligand-associated fluctuations in fluorescence (deriving from hemodynamics, pH changes, autofluorescence, motion artefact, mechanical shifts, and so on), a fit reference signal was subtracted from the dLight channel (see ‘Photometry active referencing’). Finally, referenced dLight traces were *z*-scored using a 20-s sliding window with a single sample step size slid over the entire experiment to remove slow trends in $\Delta F/F_0$ amplitudes due to long timescale effects—for example, photobleaching. Only experiments where the maximum percentage $\Delta F/F_0$

exceeded 1.5 and the dLight to reference correlation was below 0.6 were included for further analysis.

Photometry active referencing

In order to remove the effects of motion and mechanical artefacts from downstream analysis, a fit reference signal was subtracted from the demodulated dLight photometry trace as initially mentioned in ‘Photometry pre-processing’[31,54] (Extended Data Fig. 1g). First, the reference signal was low-pass filtered with a second-order Butterworth filter with a 3 Hz corner frequency. Next, to account for differences in gain or DC offset, RANSAC ordinary least squares regression was used to find the slope and bias with which to transform the reference signal to minimize the difference between the reference and the dLight photometry traces. Finally, the transformed reference trace was subtracted from the dLight trace.

Capturing 3D keypoints

To capture 3D keypoints, mice were recorded in a multi-camera open field arena with transparent floor and walls. Near-infrared video recordings at 30 Hz were obtained from six cameras (Microsoft Azure Kinect; cameras were placed above, below and at four cardinal directions). Separate deep neural networks with an HRNet architecture were trained to detect keypoints in each view (top, bottom and side) using ~1,000 hand-labelled frames[70]. Frame-labelling was crowdsourced through a commercial service (Scale AI), and included the tail tip, tail base, three points along the spine, the ankle and toe of each hind limb, the forepaws, ears, nose and implant. After detection of 2D keypoints from each camera, 3D keypoint coordinates were triangulated and then refined using GIMBAL—a model-based approach that leverages anatomical constraints and motion continuity[71]. GIMBAL requires learning an anatomical model and then applying the model to multi-camera behaviour recordings. For model fitting, we followed the approach described in ref. [71], using 50 pose states and excluding outlier poses using the EllipticEnvelope method from sklearn. For applying GIMBAL to behaviour recordings, we again followed[71], setting the parameters

obs_outlier_variance, obs_inlier_variance, and pos_dt_variance to 1e6, 10 and 10, respectively for all keypoints.

Computing 2D and 3D velocity

To compute 2D translational velocity, the centroid of the keypoints associated with the spine (approximating whole-body movement) was computed for the *x* and *y* planes (the *z* plane was disregarded). Then, the velocity was computed from the difference in position between every 2 frames and divided by 2 (to provide a smoother estimate of velocity). 3D translational velocity was computed the same way, except the *z* plane was included in the calculation. The average velocity of the keypoints associated with the forepaws were used to compute 3D forelimb velocity.

Partialing kinematic parameters from dLight

To compute the relationship between dLight and forelimb velocity, other kinematic parameters known to be correlated with dLight were partialed out of the dLight fluorescence signal. Specifically, 2D velocity, 3D velocity and height were partialed out of dLight using linear regression. Then, the correlation between the partialed dLight signal and 3D forelimb velocity were computed and compared to 1,000 bootstrapped shuffles.

Movement initiation analyses

A changepoint detection algorithm was used to find moments where mice transitioned from periods of relative stillness to movement. To capture long bouts of movement, the velocity of the 2D centroid of the mouse was *z*-scored across each experiment and then smoothed with a 50-point (1.67s) boxcar window. To find sharp changes in velocity, the derivative of smoothed velocity trace was computed, and the result was raised to the third power. Peaks in this velocity changepoint score were discovered using SciPy’s findpeaks function with the following parameters: height 1, width 1, prominence 1 so that consecutive data points around each peak were disregarded.

dLight time warping

To account for variability in syllable duration, dLight traces were time warped for Extended Data Fig. 4a. Here, all dLight traces were linearly interpolated using the numpy.interp function to a duration of 0.83 s, or 25 samples. Thus, syllables longer than 0.83 s were linearly compressed, and syllables shorter than 0.83 s were linearly expanded. We obtained similar results time warping dLight traces to 0.4 s; thus, the duration of time warped instances did not affect interpretation of subsequent analyses.

dLight average waveform *z*-scoring

For dLight waveforms shown in Fig. 1f, top and bottom, h,i,k and Extended Data Figs. 4c–g, 5c,f and 7c, first onset-aligned waveforms were *z*-scored using the mean and s.d. of fluorescence values from 10 s prior to 10 s after onset. Next, to account for differences in the number of syllable instances (trials) in each average, waveforms were additionally normalized by *z*-scoring relative to the mean and s.d. of 1,000 shuffle averages, where individual trials were circularly permuted prior to averaging.

Decoding syllable identity from dLight waveforms

To decode syllable identity from dLight waveforms or dLight peaks, a random forest classifier[72] (cuRF = 1,000 trees, max depth = 1,000, number of bins = 128, with cross-validation on 5 folds of data) was trained to predict syllable and syllable group identity on held-out data (similar to ref. [27]). Syllable groups were created by hierarchically clustering syllables based on their pairwise MoSeq distance (see below) and thresholds were increased with a distance cut-off in steps of 0.2. The inputs to the random forest classifier were either: (1) the maximum *z*-scored dLight value from syllable onset to 300 ms after syllable onset for each syllable instance or (2) dLight waveforms and their derivatives starting at syllable onset up to 300 ms into the future for individual syllable instances. Held-out accuracy was compared to 100 shuffles of syllable identity.

Decoding turning orientation from dLight waveforms

To decode turning orientation from dLight waveforms (Extended Data Fig. 5c), a linear support vector machine was trained to classify whether a particular syllable instance is a left- or rightward turning syllable using cross-validation on five folds of data. To sample the behaviour space of turning syllables, eight syllables with the largest angular velocities were chosen, four for each turning orientation. The model was fit to dLight waveforms and their derivatives starting at syllable onset up to 300 ms after onset for individual syllable instances and was tested on held-out data.

MoSeq distance

The MoSeq distance between two syllables was computed as previously described[27]. In brief, the estimated autoregressive matrices for each syllable were used to generate synthetic trajectories through principal component space (that is, in the space defined by the first ten principal components of the depth video). Then, the correlation distance between trajectories for all pairs of syllables were computed. Since the online and offline variants of MoSeq used the same autoregressive matrices, these distances are equivalent in the online and offline variants.

Analysing the relationship between dLight and syllable statistics within an experiment across syllables

The dLight fluorescence associated with syllable transitions was computed as the maximum *z*-scored dLight value from syllable onset to 300 ms after syllable onset for each syllable transition, to account for jitter in dopamine release or technical jitter in defining syllable changepoints. Throughout the text, we refer to syllable-associated waveform peak amplitudes in *z*-scored $\Delta F/F_0$ units as ‘syllable-associated dLight’. These dLight values were then averaged for each syllable and for each experiment. To assess the correlation between syllable-associated dLight and syllable counts, the dLight averages were *z*-scored across syllables in each experiment. These normalized dLight peaks represented whether a syllable had relatively higher or lower dLight during a given experiment. Finally, experiment-normalized dLight values along with syllable counts were then averaged across experiments for each mouse, thus leaving a value for each mouse and each syllable.

In order to measure the linear relationship between dLight peak values and syllable counts, a robust linear regression using the Huber regressor[73] predicted average syllable counts from average dLight peaks. The regression model was evaluated using a fivefold cross-validation repeated 100 times. Reported correlation values in Figs. 1j and 2 were estimated over the held-out data. *P*-values were estimated by comparing held-out correlation values to those estimated from a linear model computed on shuffled data. To remove syllables that varied due to finite size effects, only syllables that occurred at least 100 times total across all experiments per mouse were included.

To compute syllable entropy (estimating the randomness of outgoing transitions associated with each syllable), the outgoing transition probabilities associated with each syllable for each mouse were computed by counting the number of occurrences a syllable transitions to all others within an experiment and expressing this as a probability distribution. Next, the Shannon entropy was estimated over the outgoing transition probabilities for each syllable. Finally, the linear regression was estimated using the exact same procedure used for syllable counts.

Analysing the relationship between dLight and syllable statistics across experiments for each syllable

This series of analyses queried a total of 379 experiments. To capture the correlation between syllable-associated dLight peaks and syllable-associated behavioural features (syllable frequency, syllable entropy) within syllables but across experiments, first, the maximum *z*-scored dLight amplitude from onset to 300 ms after syllable onset at each syllable transition was computed. These syllable-associated dLight peaks were averaged for each experiment and syllable. Then, the dLight peak averages for each syllable and mouse were *z*-scored separately across experiments. Additionally, to put variation of each syllable across experiments on the same scale, syllable frequency, and syllable entropy were also *z*-scored for each syllable and mouse across experiments (Fig. 2b,i, bottom). Next, to remove variability in the calculation, values were pooled across syllables for each experiment, thus leaving a value for each experiment and mouse. To remove syllables that varied due to finite size effects, first only syllables that occurred at least 50

times per session on average were considered for downstream analysis. Linear models (Huber regressors) were fit to the resulting average dLight peaks, syllable frequency, and syllable entropy and evaluated as described in the previous section.

Analysing the moment-to-moment relationship between dLight and syllable statistics within an experiment

This series of analyses queried a total of 760 syllable–experiment pairs. dLight peak values were estimated by taking the maximum dLight value from onset to 300 ms after onset at each syllable transition. Velocity, syllable counts, and dLight peak values were averaged per syllable and per mouse over an expanding bin size; that is, velocity, syllable counts, and dLight peak values were estimated over the subsequent *n* syllables after the transition were dLight value was calculated, where *n* varied from 5 syllables up to 400 (Fig. 2e). For sequence randomness, to avoid finite size effects, dLight values were binned into 20 equally spaced bins per syllable (Fig. 2k). Then, transition matrices were combined within each bin across all syllables per mouse and per time bin. Finally, Pearson correlation values were then calculated between dLight values and the behavioural features estimated at each bin size. Pearson coefficients were *z*-scored using the mean and s.d. from Pearson coefficients estimated after shuffling dLight peak values.

Note that, in order to prevent the measurement from being influence by consistent non-stationarities in behaviour, these correlations were computed within each of the five time segments shown by dashed lines in Extended Data Fig. 2e. Then, per-segment correlations were averaged.

Time-constants associated with the correlation between dLight values and behavioural features over increasing bin sizes were estimated by fitting an exponential decay curve to the correlation values at each bin size using the SciPy's curvefit function[74]. Decay functions were fit over 1,000 bootstrap resamples of the data; the depicted distributions are taus fit over each resample.

Analysing the cross-correlation between syllable-associated dLight and syllable usage

The dLight fluorescence associated with all instances of a given syllable was binned across a three-minute window (chosen based upon the decay in Fig. 2f) and correlated with the use of that same syllable across a 3-min window, with the windows shifted the indicated amounts (*x*-axis). Correlation values (in Fig. 2g,h) were *z*-scored using the mean and s.d. from shuffles. *P*-values were estimated via shuffle test.

Analysing the relationship between syllable-associated dLight and syllable classes

Syllables were manually classified into 6 classes by hand-labelling crowd videos summarizing model output[4,27,60]. Then, syllable-associated dLight was averaged for all syllables within each class.

Encoding model predicting average dLight from behaviour

As with the linear regression analysis (previous section), dLight peaks were estimated by taking the maximum *z*-scored dLight amplitude from syllable onset to 300 ms after onset. Behavioural features (entropy, velocity, and syllable counts) after each transition were computed across various bin sizes as described in 'Analysing the relationship between dLight and syllable statistics within an experiment'. The bin sizes used were 5, 10, 25, 50, 100, 200 300, 400, 800 and 1,600 syllables. Syllable frequency, syllable entropy and velocity were averaged for each experiment and syllable in each bin size. These syllable and experiment-wide average values were then *z*-scored separately for each mouse and then averaged for each mouse and each syllable. In order to remove correlations between behavioural features they were whitened using zero-phase component analysis (ZCA) whitening. Whitened behavioural features were then fed to a Bayesian linear regression model to predict average dLight peak amplitudes per syllable and per mouse according to the following equation:

$$p(\,y|X,\beta ,{\sigma }^{2})=N({\beta }^{T}X,{\sigma }^{2})$$

where X is defined as features, β is regression coefficients, y is dLight peak values, σ is the s.d., and N is the normal distribution. A normal prior was placed on the regression coefficients, and an exponential prior was placed on

the s.d. Samples from the posterior were drawn via the no u-turn sampler (NUTS) using NumPyro (n = 1,000 warmup samples, then n = 3,000 samples)[75]. To assess the temporal relationship between behavioural features and dLight, a separate model was fit at each lag (here, features were whitened separately within each lag, Extended Data Fig. 6c). Overall model performance was quantified by feeding in features at their approximate best bin size to the model. For kinematic parameters and for entropy, this bin size (lag) was 10 timesteps; for syllable counts, this bin size (lag) was 100 timesteps (in syllable time). Then, each feature was fed in separately to quantify the performance of feature subsets.

Encoding model predicting instantaneous dLight from behaviour

In order to predict instantaneous dLight amplitudes from syllable counts, syllable entropy, velocity (2D, angular and height velocity), and acceleration, a series of convolution kernels were estimated, each of which map from each behavioural feature to dLight amplitude. Mathematically, the model can be written as follows:

$${\rm{dLight}}\left(t\right)=\sum _{f\in F}\mathop{\sum }\limits_{t{\prime} =-2s}^{2s}{{\rm{\beta }}}_{f}\left(t-{t}^{{\prime} }\right)f\left(t\right)$$

where dLight (t) corresponds to the dLight trace at time step t, $f(t)$ is the behavioural feature at time step t, and β is the weight of the convolution kernel. Kernel weights were optimized using a Huber loss via the Jax library[76]. That is to say, the dLight amplitude at each time sample is predicted by convolving each behavioural feature (frequency, entropy, velocity, and acceleration) with a convolution kernel and then summing the result across features. The model was trained and evaluated using twofold cross-validation by recording experiment, and the Pearson correlation between predicted dLight amplitudes and actual amplitudes was assessed on held-out experiments. In order to remove the effects of high frequency noise on training and evaluation, the dLight traces were smoothed using a 60-sample (2-s) boxcar filter prior to training and evaluation.

Decoding model predicting behaviour from dLight

The decoding model was designed to capture the two main effects of dopamine on behavioural statistics—usage and sequencing. The goal of the decoding model is to predict the likelihood of a sequence of syllables given past dopamine. The model comprises two key features: (1) a component that scales syllable usage by past syllable-associated dopamine, and (2) a component that scales randomness of the next syllable choice by past global dopamine. This can be summed up with the following equation:

$$P({s}_{t}=i)\propto \exp \left(\frac{{\alpha }_{a}\mathop{\sum }\limits_{n=1}^{250}\left({\rm{d}}{a}_{t-n}\exp \left(\frac{-n}{{\tau }_{a}}\right)\delta \left({s}_{t-n}=i\right)\right)}{{\alpha }_{b}\mathop{\sum }\limits_{n=1}^{250}\left({\rm{d}}{a}_{t-n}\exp \left(\frac{-n}{{\tau }_{b}}\right)\right)}\right)$$

where s_t is the syllable a mouse performs at time t during a behaviour experiment, da_t is the peak dLight recorded for syllable s_t, τ_a and τ_b describe the timescale of the usage and choice randomness component respectively, α_a and α_b scale the usage and choice randomness components respectively, and δ is the Dirac delta function (that is, one-hot encoding) that returns 1 when $s_{t-1} = i$ and 0 otherwise.

The parameters α_b, τ_a and τ_b were fixed using approximations of analysis of the behavioural data (Fig. 2), and only α_a was learned by maximizing the likelihood of the function above given the sequence of syllables mice perform across a group of experiments and peak dLight measurements associated with the syllable sequence *z*-scored across each experiment. This was done via evaluating the likelihood of the function over multiple values of α_a. τ_a (describing the effect of dopamine on future syllable usage/counts) was fixed at 100 syllable timesteps, and τ_b (describing the effect of dopamine on syllable sequence entropy) was fixed at 10 syllable timesteps. These values were approximated from the median τ values reported in Fig. 2.

To test model performance, data were split into 5 folds of training and test experiments and repeated 100 times using repeated *K*-fold cross-validation. We then computed the Pearson correlation between syllable counts from model simulations and actual syllable counts after smoothing with a 50-point rolling average. The one free parameter was fit using the training dataset and assessed on the test dataset. To avoid degradation in performance due to syllable sparsity, the top 10 syllables were used. The model was compared to a suite of control models, each evaluated over the same folds. The dopamine phase shift model was evaluated on the same data, but with all dopamine traces circularly shifted by a random integer between 1 and 1,000, and the noise model was evaluated with dopamine traces replaced by numbers drawn from a unit variance random normal distribution (since the traces were *z*-scored). In order to determine the maximum possible performance, the per experiment number of counts per syllable was correlated with the across-experiment average. Here, the model performed significantly better than controls. Median Pearson correlation between held-out predictions and observed data: actual model $r = 0.20$, phase shift control $r = 0.04$, noise model $r = 0.04$. Comparison between actual model and controls, $P = 7 \times 10^{-18}$, $U = 2,500$, $f = 1$, Mann–Whitney U test, $n = 50$ model restarts.

To test the hypothesis that endogenous and exogenous dopamine linearly combine to alter the future usage of single syllables of behaviour, the present decoding model was modified. Maximal correlations were identified between predicted and observed syllable usages when adding (or subtracting) extra dopamine (termed ‘extra DA’) to the syllable-associated dopamine amplitudes observed on catch trials (Fig. 4g–i). Model-based log likelihoods of held-out syllable choices from Opto-DA stimulation day experiments were then computed. Other versions of this model (shown in Fig. 4h) included: (1) a control model in which no ‘extra DA’ is added to the model (‘no offset’), (2) a control that uses a phase-shifted version of the dLight trace (‘random shift’), and (3) a model that uses random numbers from a normal distribution with mean and variance matched to the dLight signal (‘noise’).

dLight calibration experiments

In order to characterize the speed and magnitude of evoked dopamine transients in the open field, dLight transients were elicited using brief optogenetic stimulation of SNc axons in the DLS expressing ChrimsonR while mice freely explored an open field arena[77]. A number of stimulation parameters were tested, using varying light intensity, stimulation length, and whether the stimulus was delivered in as a single continuous-wave pulse or delivered as multiple rapid short pulses. A single, short (250 ms; roughly the timescale of syllables), continuous stimulation pulse of red light at 10 mW (Opto Engine MRL-III-635; SKU: RD-635-00500-CWM-SD-03-LED-0) most effectively matched the amplitude and dynamics of endogenous dLight transients observed in the open field. The mean Opto-DA peak was measured at $2.18 \pm 0.85\ \Delta F/F_0\ (z)$, mean spontaneous peak = $2.23 \pm 0.62\ \Delta F/F_0\ (z)$ and 99th percentile spontaneous peak = $3.40\ \Delta F/F_0\ (z)$ Pulsed stimulation was also disfavoured as numerous studies have shown that pulsed stimulation can cause synchrony in neural and axonal networks that can evoke prolonged release[78,79,80]. Note that when excited with 635 nm light, the efficiency with which light evokes spiking in neurons expressing ChrimsonR is similar to efficiency with which blue light evokes spiking in neurons expressing ChR2[77].

Once a single 250 ms continuous pulse of 10 mW light was preliminarily chosen as the desired optogenetic stimulus to evoke dopamine release from DLS dopamine axons, another round of open-loop stimulation with these stimulation parameters was performed in the open field in two of the 10 total mice injected with dLight and ChrimsonR. In these two mice, the intervals between stimulation times were drawn by randomly choosing an integer delay between 6 and 17 s for each stimulation. This range was chosen to guarantee each animal received at least 100 stimulations during an experiment. This enabled analysis of more stimulation trials with intended parameters to verify that the amplitude of evoked transients were within the same order of magnitude as spontaneously evoked transients (Fig. 3c).

DMS dLight recordings

As a series of control experiments to establish the specificity of DLS dopamine encodings, dLight recordings were performed in the DMS using the same techniques described above. dLight stereotactic injections in wild-type mice of either sex (C57BL/6J, $n = 8$) were performed at AP: 0.26, ML: 1.5, and DV: −2.2. Fibres for photometry (in C57BL/6J mice, $n = 8$, $n = 64$ recording experiments) were implanted in the manner described above at coordinates: AP: 0.26, ML: 1.5, DV: −2.0. Open field behavioural recordings and encoding models were performed for these data exactly as described above.

Stereotaxic surgery for optogenetics

Eight- to fifteen-week-old DAT-IRES-cre::Ai32 mice resulting from the cross of DAT-IRES-cre mice (The Jackson Laboratory, 006660) and Ai32 mice (The Jackson Lab, 012569) of either sex were used. The double transgenic DAT-IRES-cre::Ai32 mouse line has previously been used to conduct specific dopaminergic neuron activation[10,81,82]. Similar surgical procedures were used as described above, except two 200 µm 0.37 NA multimode optical fibres were implanted bilaterally over DLS (AP 0.260; ML 2.550; DV −2.300), in DAT-IRES-cre::Ai32 mice ($n = 20$). Control animals (DAT-IRES-cre mice, $n = 12$) of either sex were implanted bilaterally at the same coordinates, with 6 of these animals implanted in the nucleus accumbens (AP 1.300; ML 1.000; DV −4.000). These animals are collectively termed 'no-opsin controls' throughout the manuscript. Medical-grade titanium headbars were secured to the skull using cyanoacrylate. Optical stimulation experiments were then performed 2–3 weeks post-surgery.

Closed-loop stimulation behavioural paradigm

For two days prior to the closed-loop stimulation schedule (Fig. 3d), mice were habituated to the bucket for two 30-min experiments on each day. To test the change in statistics of specific syllables via syllable-triggered optogenetic stimulation, experiments were performed in a three-day schedule for each of six chosen target syllables. On the first day, two 30-min experiments were run for each mouse to characterize baseline target

syllable usage. On the second day, two 30-min 'stimulation' experiments were performed for each mouse. During these experiments blue light (470 nm, 10 mW, a single 250-ms continuous-wave pulse) was delivered on 75% of target syllable detections. Stimulation was not conditioned on syllables occurring before the target. Finally, on the third day, baseline experiment recordings were repeated to assess syllable usage memory and usage decay after reinforcement. For half of the targeted syllables for each mouse (randomized across mice), the pre-stimulation baseline experiment is the same experiment as the post-stimulation baseline experiment for a different syllable (see Fig. 3d). A three-day cadence with multiple, short behavioural recording experiments per day was chosen to both minimize non-stationarities in syllable usage within an experiment, as well as to not expose the mice to the behavioural arena for more than one total hour per day. To control for order effects on changes in target syllable usages over time, animals were randomly split into two groups, each of which had a unique ordering of target syllables across the six stimulation days of the three-week cadence. The time interval between the first experiment and the second experiment for the same mouse on each day (either recording or stimulation) was 195 min on average ±58 min (s.d.). Mice were euthanized following completion of behavioural tests, and histology was performed using procedures described above.

To assess the effect of increased dopamine release these experiments were repeated with 3-s pulsed stimulation (25 Hz, 5 ms pulse width) in $n = 3$ DAT-cre::Ai32 and $n = 2$ (DAT-IRES-cre) control animals.

Closed-loop velocity modulation experiments

DAT-IRES-cre::Ai32 mice ($n = 5$) of either sex underwent 90-min recording and manipulation experiments. For the first 30 min, we estimated the distribution of velocities for a specific target syllable. Then, for the next 30 min, optogenetic stimulation was triggered both when the syllable was expressed according to our closed-loop system and when the animal's syllable-specific velocity exceeded the 75th percentile or went below the 25th percentile. Experiments were analysed only if the mouse received at least 50 stimulations and they increased the usage of the target syllable on

average relative to their average baseline (established via separate recording experiments with no stimulation).

Quantifying changes in target syllable counts

First, the number of times the targeted syllable was performed during a 30-s sliding window (non-overlapping) for each 30-min stimulation experiment was computed. Then, a cumulative sum was taken. To turn the result into an estimate of excess target counts, a cumulative sum was also computed from the morning and evening experiments from the most recent previous baseline day. Finally, the average of the morning and evening baseline estimates was averaged and subtracted off.

'Learner' mice were defined as mice whose average change in target counts above baseline across all syllables exceed the maximum average change in target counts exhibited by no-opsin control animals. These $n = 9$ animals were used for subsequent analyses of target kinematics and learning specificity (Extended Data Fig. 10).

Quantifying effects on syllables near the target in time

To assess whether syllables temporally adjacent to the target were reinforced as a result of optogenetic stimulation, syllables were identified that—on average—were near to the target in time. Specifically, the average time between all non-targeted syllables and the target was computed, along with their change in counts above baseline. Then, syllables were binned when they occurred on average relative to the target in syllable units in equally spaced bins from ten syllables before the target to ten syllables after. Finally, for each experiment, a weighted average of the change in counts above baseline for all syllables in each bin was computed, where a syllable's weight was defined by its relative frequency in an experiment.

Quantifying effects on syllables whose velocity was similar to the target

To understand whether syllables with similar velocity profiles to the target were also reinforced, the average velocity from onset to offset for each

syllable was computed and *z*-scored across instances within an experiment. Then, the average velocity of the target was subtracted from each syllable's average velocity. Finally, the change in count above baseline for each syllable was binned by its target-velocity-difference.

Quantifying Opto-DA effects on movement parameters and sequence randomness

To quantify the effects of Opto-DA on movement parameters and sequence randomness over short timescales, sequence entropy, velocity (2D, angular and height velocity) and acceleration were estimated in five-syllable-long non-overlapping bins starting from stimulation onset. This window was chosen to minimize noise in downstream calculations while retaining reasonable time-resolution. To compensate for non-stationarities in behaviour across the experiments, mice, and targeted syllables, entropy, velocity and acceleration pre-stimulation-onset were subtracted from their values post-stimulation. Finally, these baseline-subtracted values were *z*-scored using the mean and s.d. estimated from catch trials.

Analysing the influence of dopamine on optogenetic reinforcement

Mice used to assess the influence of endogenous dopamine fluctuations on optogenetic reinforcement

As described above, eight mice injected with dLight and ChrimsonR were also run through closed-loop reinforcement experiments. The reinforcement experiment run with 250 ms 10mW CW stimulation enabled decoding analysis of how exogenous dopamine release altered usage of syllables during experiments in which 'extra DA' was added (Fig. 4g–i).

Predicting the amount of exogenously added dopamine during Opto-DA experiments using the decoding model

To predict the magnitude of exogenously evoked dLight fluorescence using the decoding model, dLight fluorescence on each instance in which the

mouse expressed the target syllable and received stimulation was replaced with the average dLight fluorescence observed for the target syllable on catch trials during which there was no optogenetic stimulation. Then an offset (denoted as ‘extra DA’) was added to each syllable instance in which the mouse received stimulation. The likelihood of the syllable sequences expressed during Opto-DA experiments was computed for a range of extra DA offsets (and hence a range of exogenously added dopamine). The model was evaluated using the exact same procedure described in ‘Decoding model predicting behaviour from dLight’, except the repeated K-fold splits (5-fold split repeated 100 times) was performed over stimulation experiments. The ‘extra DA’ outputs of the model were compared to empirical photometric data collected from animals expressing dLight that underwent ChrimsonR-mediated closed-loop reinforcement (Fig. 4i).

Using the influence of endogenous dopamine to predict Opto-DA reinforcement

In order to assess whether the impact of dopamine at baseline could predict Opto-DA reinforcement, we used the correlation between dopamine fluctuations and syllable statistics (usage and entropy) within an experiment. Specifically, we computed the correlation between dLight levels and usage as outlined in ‘Analysing the moment-to-moment relationship between dLight and syllable statistics within an experiment’ (Fig. 2e,k), except correlations were assessed per mouse and per syllable. Values at each bin size were *z*-scored using the mean and s.d. correlations computed over shuffled data. Here, n = 100 shuffles were used for the correlation with entropy for computational efficiency. To determine the modulation depth of these correlation curves for each mouse and syllable, we used the s.d. of the correlation values across bin sizes. This resulted in a value that reflected the short-term influence of dopamine on usage (Endo-DA count) and entropy (Endo-DA entropy) for all syllable–mouse pairs. Finally, these estimates were averaged per mouse for Fig. 4b,c, and per syllable for Fig. 4d. Then, the $\log_2$ fold change in target counts on stimulation days relative to baseline days was used as an estimate of Opto-DA learning. To mitigate mouse-to-mouse variability, the $\log_2$ fold change in target counts was normalized by computing the $\log_2$ fold change in target

counts against all pairs of non-stimulation days per mouse. The mean and s.d. of this distribution was used to *z*-score Opto-DA learning per mouse.

Bayesian linear regression models were used in Fig. 4b,c. A normal prior was placed on the regression coefficients, and an exponential prior on the variance. Samples from the posterior were drawn via the no u-turn sampler (NUTS) using NumPyro ($n = 1,000$ warmup samples, $n = 2,000$ samples)[75]. Performance was assessed using leave-two-out cross-validation. The linear regression model presented in Fig. 4f utilized a Huber regressor[73]. Performance of the Huber regressors was assessed using fivefold cross-validation repeated five times.

Applying RL models to open field behaviour

Reinforcement-only RL model

RL models have four key components: a reward signal, a state, a state-dependent set of available actions, and a policy (which governs how actions are chosen). Here, a simple Q-learning agent with a softmax policy was designed to model mouse behaviour in the open field as an RL process over endogenous dopamine levels[44]. Our model was recast (specifically a Q-learning agent with a softmax policy) to use endogenous dopamine (that is, syllable-associated dLight) as a reward signal, behavioural syllables as states, and transitions between behavioural syllables as actions. Given a syllable at time $t + 1$, the dLight peak occurring during the syllable at time t is considered the 'reward'. The Q-table for the model was initialized with a uniform matrix with the diagonal set to 0, since by definition there are no self-transitions in our data. For every step of each simulation, given the currently expressed syllable (that is, the state), the model samples possible future syllables (actions) based on the behavioural policy and the expected dLight transient magnitude (expected reward, specified by the Q-table) associated with each syllable transition. Then, the model selected actions according to the softmax equation

$$p(a|s)=\frac{{e}^{{Q}_{s}(a)/\tau }}{\mathop{\sum }\limits_{b=1}^{n}{e}^{{Q}_{s}(b)/\tau }}$$

where τ is the temperature. The model is fed 30-min experiments of actual data. Data was formatted as a sequence of states and syllable-associated dopamine. Given the current state, the model selects an action according to the softmax equation. To update the Q-table and simulate the effect of endogenous dopamine as reward, the syllable-associated dopamine is presented to the model as reward in a standard Q-learning equation. Specifically, the Q-table was then updated according to

$$Q({s}_{t},{a}_{t})\leftarrow Q({s}_{t},{a}_{t})+\alpha [{r}_{t+1}+\gamma {max}_{a}Q({s}_{t+1},a)-Q({s}_{t},{a}_{t})]$$

where Q is the Q-table that defines the probability of action a while in state s, α is the learning rate, r is the reward associated with action a and state s (the dLight peak value at the transition between syllable a and syllable s), and γ is the discount factor. Performance was assessed by taking the Pearson correlation between the model's resulting Q-table at the end of the simulation and the empirical transition matrix observed in the experimental data. Here, each row of the empirical transition matrix and the Q-table were separately z-scored prior to computing the Pearson correlation. Note that the learned Q-table is functionally equivalent to a transition matrix in this formulation. To avoid degradation in performance due to syllable sparsity, the top 10 syllables were used.

Dynamic RL model

To account for the short-term effect of dopamine on sequence randomness, a dopamine-dependent term was added to the baseline model's policy

$$p(a|s)=\frac{{e}^{{Q}_{s}(a)/\tau (t)}}{\mathop{\sum }\limits_{b=1}^{n}{e}^{{Q}_{s}(b)/\tau ({\rm{t}})}}$$

where temperature is now time-dependent and evolves according to,

$$\tau \left(t\right)=I\left(t\right)\exp \left(\frac{t-n}{{{\tau }}_{{\rm{decay}}}}\right)+\,{\tau }_{{\rm{baseline}}}$$

and,

$$I\left(t\right)=v\,\,\text{if}\,\,r\left(t\right)\ge \lambda $$

Here, τ_{decay} corresponds to the time constant with which dopamine's effect on temperature decays, $\tau_{baseline}$ is the baseline temperature, v is the amount by which temperature is increased if the $r(t)$ goes above the threshold λ, and n is the number of timesteps after the threshold has been crossed. Experiments were split into training and test datasets via twofold cross-validation, and the training set was used to fit all free parameters. To compare the dynamic to the reinforcement-only model, v was set to 0—this turns off the temperature varying component of the dynamic model. Note that we observe qualitatively similar results under an alternative formulation. Rather than feeding the model 30-min sessions of actual data, we allow the model to freely select actions, and reward was randomly drawn from dLight peaks associated with that action in actual data.

Reward-prediction error model variant

Models were fit using observed dopamine magnitude as either the (1) reward term (see above) or (2) reward-prediction error term $[{r}_{t+1}+\gamma {max}_{a}Q({s}_{t+1},a)-Q({s}_{t},{a}_{t})]$. For each model type, a grid search was performed across values of α (learning rate), γ (discount factor, used in the reward model only), and temperature (randomness of the next action). Held-out log likelihood was computed for each fit and z-scored using the mean and variance of the held-out log likelihood from models fit to data shuffled between experiments (n = 10 shuffles). This comparison is only valid for our particular model formulation. There are alternative formulations for which dopamine acting as a reward-prediction error are consistent with our data.

Statistics

All hypothesis tests were non-parametric. Effect sizes for Mann–Whitney U tests are presented as the common language effect size f. Correlations were established as significant by comparing to n = 1,000 shuffled correlations (referred to as the shuffle test throughout the manuscript). For shuffle test if all correlations exceeded the 1,000 shuffles, the P-value is listed as P <

0.001 rather than $P = 0$. P-values were adjusted to account for multiple comparisons where appropriate using the Holm–Bonferonni stepdown procedure. Sample sizes were not pre-determined but are consistent with sample sizes typically used in the field. For examples using similar techniques see[10,14]. Blinding was not performed, but MoSeq-based analysis of behaviour was automated.

Plotting

Box plots (here and throughout) obey standard conventions: edges represent the first and third quartiles, whereas whiskers extend to include the furthest data point within 1.5 interquartile ranges of either the first or third quartile.

Software packages

In addition to analysis-specific packages cited in the relevant sections above, the following packages were used for analysis: NumPy[83], Python[84], Seaborn[85], Matplotlib[86] and Python 3 (ref. [87]).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data that support the findings of the current study are available on Zenodo at https://doi.org/10.5281/zenodo.7274803.

Code availability

All code related to this study was developed in MATLAB or Python and is available at http://github.com/dattalab/dopamine-reinforces-spontaneous-behavior.

References

1. Tinbergen, N. *The Study of Instinct* (Clarenden Press, 1951).

2. Berridge, K. C., Fentress, J. C. & Parr, H. Natural syntax rules control action sequence of rats. *Behav. Brain Res.* **23**, 59–68 (1987).

3. Gray, J. M., Hill, J. J. & Bargmann, C. I. A circuit for navigation in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **102**, 3184–3191 (2005).

4. Wiltschko, A. B. et al. Mapping sub-second structure in mouse behavior. *Neuron* **88**, 1121–1135 (2015).

5. Johnson, R. E. et al. Probabilistic models of larval zebrafish behavior reveal structure on many scales. *Curr. Biol.* **30**, 70–82.e74 (2020).

6. Panigrahi, B. et al. Dopamine is required for the neural representation and control of movement vigor. *Cell* **162**, 1418–1430 (2015).

7. Berridge, K. C. Substantia nigra 6-OHDA lesions mimic striatopallidal disruption of syntactic grooming chains: a neural systems analysis of sequence control. *Psychobiology* **17**, 377–385 (1989).

8. Benecke, R., Rothwell, J. C., Dick, J. P., Day, B. L. & Marsden, C. D. Disturbance of sequential movements in patients with Parkinson's disease. *Brain* **110**, 361–379 (1987).

9. da Silva, J. A., Tecuapetla, F., Paixao, V. & Costa, R. M. Dopamine neuron activity before action initiation gates and invigorates future movements. *Nature* **554**, 244–248 (2018).

10. Coddington, L. T. & Dudman, J. T. The timing of action determines reward prediction signals in identified midbrain dopamine neurons. *Nat. Neurosci.* **21**, 1563–1573 (2018).

11. Dodson, P. D. et al. Representation of spontaneous movement by dopaminergic neurons is cell-type selective and disrupted in

parkinsonism. *Proc. Natl Acad. Sci. USA* **113**, E2180–E2188 (2016).

12. Barter, J. W. et al. Beyond reward prediction errors: the role of dopamine in movement kinematics. *Front. Integr. Neurosci.* **9**, 39 (2015).

13. DeLong, M. R., Crutcher, M. D. & Georgopoulos, A. P. Relations between movement and single cell discharge in the substantia nigra of the behaving monkey. *J. Neurosci.* **3**, 1599–1606 (1983).

14. Howe, M. W. & Dombeck, D. A. Rapid signalling in distinct dopaminergic axons during locomotion and reward. *Nature* **535**, 505–510 (2016).

15. Puryear, C. B., Kim, M. J. & Mizumori, S. J. Y. Conjunctive encoding of movement and reward by ventral tegmental area neurons in the freely navigating rodent. *Behav. Neurosci.* **124**, 234–247 (2010).

16. Engelhard, B. et al. Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons. *Nature* **570**, 509–513 (2019).

17. Palmiter, R. D. Dopamine signaling in the dorsal striatum is essential for motivated behaviors: lessons from dopamine-deficient mice. *Ann. NY Acad. Sci.* **1129**, 35 (2008).

18. Niv, Y. Cost, benefit, tonic, phasic. *Ann. NY Acad. Sci.* **1104**, 357–376 (2007).

19. Hughes, R. N. et al. Ventral tegmental dopamine neurons control the impulse vector during motivated behavior. *Curr. Biol.* **30**, 2681–2694.e2685 (2020).

20. Saunders, B. T., Richard, J. M., Margolis, E. B. & Janak, P. H. Dopamine neurons create Pavlovian conditioned stimuli with circuit-defined motivational properties. *Nat. Neurosci.* **21**, 1072–1083 (2018).

21. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).

22. Kim, H. R. et al. A unified framework for dopamine signals across timescales. *Cell* **183**, 1600–1616.e1625 (2020).

23. Howard, C. D., Li, H., Geddes, C. E. & Jin, X. Dynamic nigrostriatal dopamine biases action selection. *Neuron* **93**, 1436–1450.e1438 (2017).

24. Tsutsui-Kimura, I. et al. Distinct temporal difference error signals in dopamine axons in three regions of the striatum in a decision-making task. *eLife* **9**, 3327–3339 (2020).

25. Parker, N. F. et al. Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. *Nat. Neurosci.* **19**, 845–854 (2016).

26. Schwarz, R. F., Branicky, R., Grundy, L. J., Schafer, W. R. & Brown, A. E. X. Changes in postural syntax characterize sensory modulation and natural variation of *C. elegans* locomotion. *PLoS Comput. Sci.* **11**, e1004322 (2015).

27. Markowitz, J. E. et al. The striatum organizes 3D behavior via moment-to-moment action selection. *Cell* **174**, 44–58.e17 (2018).

28. Klaus, A. et al. The spatiotemporal organization of the striatum encodes action space. *Neuron* **95**, 1171–1180.e1177 (2017).

29. Shin, J. H., Song, M., Paik, S.-B. & Jung, M. W. Spatial organization of functional clusters representing reward and movement information in the striatal direct and indirect pathways. *Proc. Natl Acad. Sci. USA* **117**, 27004–27015 (2020).

30. Shen, W., Flajolet, M., Greengard, P. & Surmeier, D. J. Dichotomous dopaminergic control of striatal synaptic plasticity. *Science* **321**, 848–851 (2008).

31. Patriarchi, T. et al. Ultrafast neuronal imaging of dopamine dynamics with designed genetically encoded sensors. *Science* **360**, eaat4422 (2018).

32. Giros, B., Jaber, M., Jones, S. R., Wightman, R. M. & Caron, M. G. Hyperlocomotion and indifference to cocaine and amphetamine in mice lacking the dopamine transporter. *Nature* **379**, 606–612 (1996).

33. Hamilos, A. E. et al. Slowly evolving dopaminergic activity modulates the moment-to-moment probability of reward-related self-timed movements. *eLife* **10**, e62583 (2021).

34. Tecuapetla, F., Matias, S., Dugue, G. P., Mainen, Z. F. & Costa, R. M. Balanced activity in basal ganglia projection pathways is critical for contraversive movements. *Nat. Commun.* **5**, 4315 (2014).

35. Moss, M. M., Zatka-Haas, P., Harris, K. D., Carandini, M. & Lak, A. Dopamine axons in dorsal striatum encode contralateral visual stimuli and choices. *J. Neurosci.* **41**, 7197–7205 (2021).

36. Pesek-Cotton, E. F., Johnson, J. E. & Newland, M. C. Reinforcing behavioral variability: an analysis of dopamine-receptor subtypes and intermittent reinforcement. *Pharmacol. Biochem. Behav.* **97**, 551–559 (2011).

37. Oades, R. D. The role of noradrenaline in tuning and dopamine in switching between signals in the CNS. *Neurosci. Biobehav. Rev.* **9**, 261–282 (1985).

38. Bercken, J. H. L. V. D. & Cools, A. R. Role of the neostriatum in the initiation, continuation and termination of behavior. *Appl. Neurophysiol.* **42**, 106–108 (1979).

39. Nourizonoz, A. et al. EthoLoop: automated closed-loop neuroethology in naturalistic environments. *Nat. Methods* **17**, 1052–1059 (2020).

40. Xiao, L. et al. A basal ganglia circuit sufficient to guide birdsong learning. *Neuron* **98**, 208–221.e205 (2018).

41. Yttri, E. A. & Dudman, J. T. Opponent and bidirectional control of movement velocity in the basal ganglia. *Nature* **533**, 402–406 (2016).

42. Hamid, A. A. et al. Mesolimbic dopamine signals the value of work. *Nat. Neurosci.* **19**, 117–126 (2016).

43. Bova, A. et al. Precisely-timed dopamine signals establish distinct kinematic representations of skilled movements. *eLife* **9**, e61591 (2020).

44. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* 2nd edn (MIT Press, 2018).

45. Watabe-Uchida, M., Zhu, L., Ogawa, S. K., Vamanrao, A. & Uchida, N. Whole-brain mapping of direct inputs to midbrain dopamine neurons. *Neuron* **74**, 858–873 (2012).

46. Lahiri, A. K. & Bevan, M. D. Dopaminergic transmission rapidly and persistently enhances excitability of D1 receptor-expressing striatal projection neurons. *Neuron* **106**, 277–290.e6 (2020).

47. Lindsey, J. & Litwin-Kumar, A. Action-modulated midbrain dopamine activity arises from distributed control policies. Preprint at https://doi.org/10.48550/arXiv.2207.00636 (2022).

48. Greenstreet, F. et al. Action prediction error: a value-free dopaminergic teaching signal that drives stable learning. Preprint at *bioRxiv* https://doi.org/10.1101/2022.09.12.507572 (2022).

49. Duffy, A., Latimer, K. W., Goldberg, J. H., Fairhall, A. L. & Gadagkar, V. Dopamine neurons evaluate natural fluctuations in performance quality. *Cell Rep.* **38**, 110574 (2022).

50. Gadagkar, V. et al. Dopamine neurons encode performance error in singing birds. *Science* **354**, 1278–1282 (2016).

51. Foo, C. et al. Reinforcement learning links spontaneous cortical dopamine impulses to reward. *Curr. Biol.* **31**, 4111–4119.e4 (2021).

52. Wei, W., Mohebi, A. & Berke, J. D. A Spectrum of Time Horizons for Dopamine Signals. Preprint at *bioRxiv*

https://doi.org/10.1101/2021.10.31.466705 (2022).

53. van Elzelingen, W. et al. Striatal dopamine signals are region specific and temporally stable across action-sequence habit formation. *Curr. Biol.* **32**, 1163–1174.e6 (2022).

54. Lerner, T. N. et al. Intact-brain analyses reveal distinct information carried by SNc dopamine subcircuits. *Cell* **162**, 635–647 (2015).

55. Cotzias, G. C., Papavasiliou, P. S. & Gellene, R. Modification of Parkinsonism—chronic treatment with L-dopa. *N. Engl. J. Med.* **280**, 337–345 (1969).

56. Shettleworth, S. J. & Juergensen, M. R. Reinforcement and the organization of behavior in golden hamsters: brain stimulation reinforcement for seven action patterns. *J. Exp. Psychol.* **6**, 352–375 (1980).

57. Hintiryan, H. et al. The mouse cortico-striatal projectome. *Nat. Neurosci.* **19**, 1100–1114 (2016).

58. Hunnicutt, B. J. et al. A comprehensive excitatory input map of the striatum reveals novel functional organization. *eLife* **5**, e19103 (2016).

59. Yagishita, S. et al. A critical time window for dopamine actions on the structural plasticity of dendritic spines. *Science* **345**, 1616–1620 (2014).

60. Wiltschko, A. B. et al. Revealing the structure of pharmacobehavioral space through motion sequencing. *Nat. Neurosci.* **23**, 1433–1443 (2020).

61. Abadi, M. et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems. Preprint at https://doi.org/10.48550/arXiv.1603.04467 (2016).

62. Howe, M. W., Tierney, P. L., Sandberg, S. G., Phillips, P. E. M. & Graybiel, A. M. Prolonged dopamine signalling in striatum signals

proximity and value of distant rewards. *Nature* **500**, 575–579 (2013).

63. Robinson, D. L., Heien, M. L. A. V. & Wightman, R. M. Frequency of dopamine concentration transients increases in dorsal and ventral striatum of male rats during introduction of conspecifics. *J. Neurosci.* **22**, 10477–10486 (2002).

64. Shnitko, T. A. & Robinson, D. L. Regional variation in phasic dopamine release during alcohol and sucrose self-administration in rats. *ACS Chem. Neurosci.* **6**, 147–154 (2015).

65. Akerboom, J. et al. Optimization of a GCaMP calcium indicator for neural activity imaging. *J. Neurosci.* **32**, 13819–13840 (2012).

66. Tian, L. et al. Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nat. Methods* **6**, 875–881 (2009).

67. Barnett, L. M., Hughes, T. E. & Drobizhev, M. Deciphering the molecular mechanism responsible for GCaMP6m's Ca^{2+}-dependent change in fluorescence. *PLoS ONE* **12**, e0170934 (2017).

68. Díaz-García, C. M. et al. The distinct roles of calcium in rapid control of neuronal glycolysis and the tricarboxylic acid cycle. *eLife* **10**, e64821 (2021).

69. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).

70. Yu, C. et al. Lite-HRNet: A Lightweight High-Resolution Network. Preprint at https://doi.org/10.48550/arXiv.2104.06403 (2021).

71. Zhang, L., Dunn, T., Marshall, J., Olveczky, B. & Linderman, S. Animal pose estimation from video data with a hierarchical von Mises-Fisher-Gaussian model. In *Proc. 24th International Conference on Artificial Intelligence and Statistics.* Vol. 130, 2800–2808 (PMLR, 2021).

72. Rapids Development Team. RAPIDS: collection of libraries for end to end GPU data science. https://rapids.ai/citations.html (2018).

73. Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

74. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

75. Phan, D., Pradhan, N. & Jankowiak, M. Composable effects for flexible and accelerated probabilistic programming in NumPyro. Preprint at https://doi.org/10.48550/arXiv.1912.11554 (2019).

76. Bradbury, J. et al. JAX: composable transformations of Python+NumPy programs v. 0.3.13. https://github.com/google/jax (2018).

77. Klapoetke, N. C. et al. Independent optical excitation of distinct neural populations. *Nat. Methods* **11**, 338–346 (2014).

78. Kravitz, A. V. & Bonci, A. Optogenetics, physiology, and emotions. *Front. Behav. Neurosci.* **7**, 169 (2013).

79. Kravitz, A. V. et al. Regulation of parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry. *Nature* **466**, 622–626 (2010).

80. Histed, M. H. & Maunsell, J. H. R. Cortical neural populations can guide behavior by integrating inputs linearly, independent of synchrony. *Proc. Natl Acad. Sci. USA* **111**, E178–E187 (2014).

81. Madisen, L. et al. A toolbox of Cre-dependent optogenetic transgenic mice for light-induced activation and silencing. *Nat. Neurosci.* **15**, 793–802 (2012).

82. Tritsch, N. X., Oh, W.-J., Gu, C. & Sabatini, B. L. Midbrain dopamine neurons sustain inhibitory transmission using plasma membrane uptake of GABA, not synthesis. *eLife* **3**, 1185–1120 (2014).

83. Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).

84. McKinney, W. et al. Data structures for statistical computing in Python. In *Proc 9th Python in Science Conference* (eds van der Walt, S. & Millman, J.) 51–56 (SciPy, 2010).

85. Waskom, M. Seaborn: statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).

86. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

87. Van Rossum, G. & Drake, F. L. *Python 3 Reference Manual* (CreateSpace, 2009).

88. Arbuthnott, G. W. & Crow, T. J. Relation of contraversive turning to unilateral release of dopamine from the nigrostriatal pathway in rats. *Exp. Neurol.* **30**, 484–491 (1971).

Acknowledgements

S.R.D. is supported by NIH grants RF1AG073625 and R01NS114020, and the Brain Research Foundation. S.R.D., B.L.S. and S.W.L are supported by NIH grant U19NS113201 and and the Simons Collaboration on the Global Brain. J.E.M. is supported by a Career Award at the Scientific Interface (CASI) from the Burroughs Wellcome Fund. W.F.G. is supported by NIH grant F31NS113385. M.J. is supported by NIH grant F31NS122155. C.W. is a Fellow of the Jane Coffin Childs Memorial Fund for Medical Research. S.W.L is supported by the Alfred P. Sloan Foundation. We thank N. Bhagat, J. Araki and J. Wallace for administrative support; the Harvard Neurobiology Imaging Core for imaging support, which is supported as part of NINDS grant NS072030; the HMS Research Instrumentation Core, which is supported by the Bertarelli Program in Translational Neuroscience and Neuroengineering, and by NEI grant EY012196; S. Gershman, J. Assad, S. Shea and members of the Datta laboratory for useful comments on the paper; and S Knemeyer for mouse drawings. Portions of this research

were conducted on the O2 High Performance Compute Cluster at Harvard Medical School. Brain diagrams were created with BioRender.com. We apologize to those whose work we were unable to cite due to length limits.

Author information

Author notes

1. Jeffrey E. Markowitz

   Present address: Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA, USA

2. These authors contributed equally: Jeffrey E. Markowitz, Winthrop F. Gillis, Maya Jay

Authors and Affiliations

1. Department of Neurobiology, Harvard Medical School, Boston, MA, USA

   Jeffrey E. Markowitz, Winthrop F. Gillis, Maya Jay, Jeffrey Wood, Ryley W. Harris, Robert Cieszkowski, Rebecca Scott, David Brann, Dorothy Koveal, Tomasz Kula, Caleb Weinreb, Mohammed Abdal Monium Osman, Bernardo L. Sabatini & Sandeep Robert Datta

2. Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA, USA

   Sandra Romero Pinto & Naoshige Uchida

3. Center for Brain Science, Harvard University, Cambridge, MA, USA

   Sandra Romero Pinto & Naoshige Uchida

4. Wu Tsai Neurosciences Institute, Stanford University, Stanford, CA, USA

Scott W. Linderman

5. Department of Statistics, Stanford University, Stanford, CA, USA

   Scott W. Linderman

6. Howard Hughes Medical Institute, Chevy Chase, MD, USA

   Bernardo L. Sabatini

Contributions

J.E.M., W.F.G., M.J., B.L.S. and S.R.D. designed experiments and interpreted the data. J.E.M., W.F.G. and M.J. performed experiments and data analyses. J.W., R.W.H., R.C. and R.S. conducted behavioural experiments and histology. D.B. assisted with animal husbandry. C.W., M.A.M.O., S.R.P., N.U. and S.W.L. assisted with computational modelling. T.K. and D.K. performed HEK cell experiments. J.E.M., W.F.G., M.J. and S.R.D. wrote the manuscript with input from all authors.

Corresponding author

Correspondence to Sandeep Robert Datta.

Ethics declarations

Competing interests

S.R.D. sits on the scientific advisory boards of Neumora and Gilgamesh Therapeutics, which have licensed or sub-licensed the MoSeq technology.

Peer review

Peer review information

Nature thanks Jesse Goldberg, Genela Morris and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 dLight validation, photometry setup and motion artifact removal.

a) Maximum projection epifluorescence images of HEK cells transfected with the dLight plasmid, and excited with either 480 nm blue (top) or 400 nm UV (bottom) light. Green emission (527 nm) was collected for both excitation wavelengths. Scale bar indicates 20 µm. Three separate experiments were performed with n = 28, n = 45, n = 42 regions of interest (ROIs) from the first, second, and third experiment respectively. **b**) Scatter plot of pixel fluorescence for each pixel location under both 400 nm and 480 nm excitation. A regression line was fit to the scatter plot (blue). The strength of the correlation demonstrates that both excitation wavelengths cause similar spatial patterns of dLight emission at 527 nm. **c**) Single cell dLight responses to perfused dopamine at 480 nm (left) or 400 nm (right) wavelengths. White dashed lines indicate time points where dopamine is washed into and out of the sample. Each row is an individual cell region of interest (ROI). **d**) Correlation in single cell dLight responses to perfused dopamine imaged with blue (x-axis) or UV light (y-axis). Each dot is an individual cell. Blue line indicates linear fit. The near-zero slope indicates that the UV-light-dependent green emission is almost entirely independent of dopamine concentration. **e**) Validation of dLight1.1 response to optogenetic stimulation. To assess how quickly DLS dLight reports dopamine transients in vivo, SNc axons were optogenetically stimulated using ChrimsonR (whose excitation spectrum is separated from that of dLight) while dLight fluorescence was recorded. Left: schematic illustrating

viral injection and implant procedure to simultaneously record dLight transients and optogenetically stimulate dopamine axons in DAT-Cre animals. dLight was injected into the right dorsolateral striatum, Cre-dependent ChrimsonR was injected into the right SNc, and an optical fiber was placed above the dLight injection site in the DLS (see Methods). Middle: coronal slice depicting the expression of both dLight and ChrimsonR in the striatum. Right: mean stimulation-evoked dLight fluorescence using ChrimsonR in dopaminergic axons originating from SNc. The green shaded region indicates bootstrap SEM. The gray shaded area indicates the 95% bootstrap confidence interval of the mean trace pre-stimulation. Red shading indicates the duration of optogenetic stimulation. dLight transients are resolvable starting 67 ms from stimulation onset, suggesting that dLight can report rapid dopamine dynamics via photometry in DLS ($p = 0.005$, $U = 2610$, $f = 0.32$, one-sided Mann-Whitney U test). **f**) Schematic describing the photometry recording setup. A blue and UV LED were modulated at different frequencies, and light delivered to the mouse via a single fiber optic cable. Green fluorescence was acquired using a photodetector and the blue (dLight) and UV (isosbestic) components were separated using lock-in amplification. **g**) Example data depicting a dLight trace along with the simultaneously recorded reference signal acquired during free behaviour. The UV, or reference signal, represents the contributions of motion and mechanical artifacts and other non-ligand dependent changes in sensor fluorescence. Since UV and blue excitation of dLight cannot be perfectly matched, a gain and bias term were fit to match the UV-excited emission to the blue-excited emission per previously published papers[54]; this fitting process maximizes the correlation between the UV and blue signals, enabling us to effectively subtract the reference signal from the dLight signal. Left: example fit showing the UV trace scaled to match the dLight trace (bottom) aligned to syllables (top). Right: histogram of r^2 values between the photometry reference (UV component) and signal (dLight component), prior to fitting and subtracting the reference signal (top) and after fitting and subtracting the reference signal (bottom). Note that the baseline correlation between the reference and signal channels prior to subtraction is low. **h**) Top: the probability, on average, of observing a dopamine transient across all mice and experiments – defined as the $\Delta F/F0$ trace crossing 1 standard deviation (computed per experiment) – at a given timepoint in the experiment. The probability was estimated in one

minute bins. Shading indicates the 95% bootstrap confidence interval across per-mouse averages. Bottom: the maximum ΔF/F0 per 1 min time bin across all mice and experiments. As in the top plot, the average across all mice and experiments is shown, with shading indicating the 95% bootstrap confidence interval across per-mouse averages.

Extended Data Fig. 2 MoSeq captures subsecond structure in spontaneous mouse behaviour.

a) Distribution of syllable durations identified by MoSeq. The mean/median syllable duration was 566/400 ms +/− 636 ms SD. **b**) The average number of times each MoSeq-identified syllable is used during a 30-minute experiment per mouse (n = 16 mice). Error bars indicate bootstrap 95% confidence intervals across mice. **c**) Human-annotated descriptions of observed behavioural syllables. Left: semantic labels and “spinograms” of all behavioural syllables used more than 1% of the time, here to provide an illustration of movements associated with syllables. Each trace in the spinogram is an average height profile of the mouse computed by taking the pixel values along the center of the depth image across columns (note that MoSeq pre-processes depth images so that mice always face to the right of the cropped depth image). Each trace from left to right is the average of each frame of the behavioural syllable from the beginning to end. The distances between successive traces are proportional to the average x/y displacement from one frame to the next. Spinograms are color-coded by the average angular velocity of the syllable. Right: dendrogram computed using the pairwise MoSeq distance of all behavioural syllables (see Methods for a description of how MoSeq distances, which capture the average three-dimensional pose dynamics of each syllable, are computed). Spinograms are aligned to their corresponding leaf in the dendrogram. **d**) The average transition matrix visualized as a state map. Each circle corresponds to a syllable, and each arrow corresponds to the likelihood that there is a transition from one syllable to the next. Arrow width indicates transition probability. All transitions with a probability below 0.1 are removed for visual clarity in this statemap and in all subsequent statemaps. **e**) Kinematic parameters over time averaged across all experiments and mice, demonstrating non-stationarities in kinematics across each recording experiment. Lines indicate boundaries derived via k-means clustering of

this data. Note that these boundaries were used for analysis shown in Fig. 2e and k. Specifically, in order to prevent non-stationarities from impacting within-experiment correlations, correlations were computed within each of these segments and then averaged. **f**) Heatmap of syllable counts computed over a six-minute sliding window for the 37 syllables used >1% of the time in an example experiment. Syllables are sorted by total usage in the experiment, with the most-used syllable at the top and least used on the bottom. The colors above each segment of the plot indicate the time intervals used to compute the transition matrices in Extended Data Fig. 2g. **g**) State maps computed for each colored section of the example experiment shown in Extended Data Fig. 2f, summarizing the transition statistics between behavioural syllables, and demonstrating that transitions are also non-stationary over each imaging experiment. Each node is a syllable, and each line represents the transition from one syllable to the next (whose width specifies the observed likelihood of each transition, per the legend).

Extended Data Fig. 3 Validating correlations to kinematic variables through multi-camera keypoint tracking.

a) 15 keypoints were tracked in 3D using 6 infrared cameras (Azure Kinect, see Methods) positioned around the open field arena (n = 5 mice). A custom keypoint detection network was trained to identify all keypoints using manually labeled frames from each individual camera and integrated post-hoc with GIMBAL (see Methods). Top: schematic of keypoints positioned on the mouse. Bottom: example aligned frames from each of the 6 cameras with keypoints superimposed. **b**) Pearson correlation between 2D velocity and simultaneously recorded dLight at different timescales. 2D velocity was estimated by computing the centroid of the spinal keypoints in the X and Y plane for each frame (shown as the dark blue keypoints in Extended Data Fig. 3a) and taking the difference between centroid positions across frames. As in Fig. 1e, 2D velocity is negatively correlated with dLight transient rates at short timescales, and positively correlated at long timescales. Top: Pearson correlations between 2D velocity and dLight transient rates across various time bins. Bottom: Pearson correlations between 2D velocity and average dLight fluorescence across various time bins. Shading represents 68% CI. **c**) Left: correlation between 2D velocity and dLight fluorescence after binning the data into 400ms time points (Pearson $r = -0.16$, $p < .001$

one-sided shuffle test). Each dot is a single 400ms time bin. Color represents point density, where brighter colors indicate denser points. Right: correlation between 3D forelimb velocity and dLight fluorescence after partialing out the relationship between dLight and other known kinematic parameters such as velocity and height (see Methods, Pearson $r = -0.02$, $p < .001$ one-sided shuffle test).

Extended Data Fig. 4 Variability of syllable-specific dLight waveforms across mice and experiments.

a) Average dLight fluorescence aligned to syllable onset after time warping (traces were warped using linear interpolation from syllable onset to syllable offset, see Methods). **b**) Average probability of a syllable transition occurring near a dLight peak across all experiments. Peaks in the dLight trace were identified by first computing and z-scoring the derivative of the ΔF/F0 trace, and identifying peaks as values that exceeded the 90^{th} (left), 95^{th} (middle), or 99^{th} (right) percentile. We then plotted the probability that a syllable transition occurred given a dopamine peak. Gray shading indicates 95% bootstrap confidence interval of the shuffle. For the 95^{th} percentile threshold, a syllable transition is likeliest to occur 200 ms prior to the dLight peak. However, the estimated dLight peak lags actual dopamine release by 10s of milliseconds (Extended Data Fig. 1e)[31]. **c**) Left: schematic showing forms of variability in dLight fluorescence measurements across experiments. dLight fluorescence assessed via photometry often exhibits baseline shifts and shifts in fluorescence scaling that can be normalized across experiments by z-scoring the fluorescence trace. Z-scoring dLight per experiment will have the effect of shifting the distribution leftward (and thus producing negative values). Top right: distribution of all syllable-associated dLight peaks across all mice and all experiments (left), and corresponding cumulative distribution (right). Bottom right: distribution of all syllable-associated dLight peaks across all mice and all experiments after z-scoring fluorescence traces from each experiment (left), and corresponding cumulative distribution (right). **d**) Left: assessing variability of the average dLight transient amplitudes from mouse to mouse. Shown is the average dLight amplitude aligned to syllable transitions Z-scored relative to a shuffle as in Fig. 1f. The thick black line indicates the average

across all mice, and per-mouse averages are shown as thin gray lines. Right: same as left except averages across experiments are shown; the thick black line indicates the overall average, and the thin gray lines are per-experiment averages. **e**) Pseudo-color plots where each row depicts per-experiment average aligned to syllable onset for all experiments, grouped by mouse. Gray lines indicate boundaries between individual mice. **f**) Left: pseudo-color plot of all per-syllable dLight waveforms as in Fig. 1h, except shown for each mouse. The color bar on the left indicates which rows correspond to which syllables using the same sorting as Fig. 1h. Within a syllable-specific block, individual rows correspond to per-mouse average dLight waveforms (n = 518 syllable/mouse pairs). Average dLight waveforms are z-scored to a shuffle. Right: the syllable-associated peak dLight value for each row, computed from each waveform between 0–300 ms from syllable onset. **g**) Average dLight fluorescence aligned to syllable onset for three example syllables shown in Extended Data Fig. 4e; the thick black line indicates the mean across all experiments, with the thin grey lines indicating averages from each mouse. **h**) The probability of observing a syllable-specific dLight peak value across every syllable instance and across all mice and experiments. Syllable-specific peak values are computed using the maximum value in a 300 ms window after syllable onset. Here, color values indicate the likelihood of observing a specific peak dLight amplitude from trial to trial without averaging. Here, dLight is z-scored within each experiment. Cyan bars show the location of the overall average for each syllable. Rows are sorted in the same order as Fig. 1h and Extended Data Fig. 4f. **i**) The probability of syllable-specific average dLight peak values across experiments. Syllable-specific peak values are computed using the maximum value in a 300 ms window after syllable onset and averaged over the experiment, and thus do not correspond to the average waveform peaks in Extended Data Fig. 4e. Each row corresponds to a given syllable, and color values indicate the likelihood of observing a given peak dLight amplitude, on average, across experiments. Here, average dLight peaks are z-scored within each experiment. Cyan bars show the location of the overall average for each syllable. Rows are sorted in the same order as Fig. 1h and Extended Data Fig. 4f.

Extended Data Fig. 5 Querying different possible sources of variability in dLight waveforms.

a) Average per-syllable dLight peaks associated with six behavioural categories (n = 7 dive syllables, 7 grooms, 9 pauses, 13 rears, 5 scrunches, 16 walks). Each category is associated with multiple syllables and were identified through human annotation. **b**) Within-syllable changes in kinematic parameters do not covary with peak dLight. Box plots of kinematic parameters binned by syllable-associated peak dLight – shown are the first and fourth quartiles. Kinematic variables were averaged from syllable onset to syllable offset, and box plots show the distribution of per-instance averages. Box plots for two examples syllables are shown, an investigatory pause (top; N = 15245 syllable instances) and a scrunch (bottom; N = 11838 syllable instances). **c**) Left top: average dLight fluorescence waveforms for two syllables that contain a left- (contralateral) and right-ward (ipsilateral) turning component. Consistent with prior studies indicating that elevated dopamine and striatal activity is associated with contralateral turning, we find higher average dLight levels are associated with contralateral turning[34,88]. Fluorescence traces were z-scored to a circular shuffle. Left bottom: dLight waveforms broken out into quartiles based on syllable-associated fluorescence, as in Fig. 1k. Right: performance of a linear SVM classifier predicting individual syllable instances as left or right turns. Average observed accuracy was 51%, indicating substantial instance-by-instance variability ($p < .001$, one-sided shuffle test). **d**) Schematic illustrating the hypothesis that dopamine fluctuations may reflect performance prediction errors; here “performance error” is defined as the degree to which a given syllable instance differs from its mean implementation (see Methods). **e**) Top: schematic describing the linear model used to characterize whether syllable rendition quality compared to a template provides additional information about dLight fluorescence on top of the kinematic parameters described in Fig. 1e. Bottom: model coefficient for each kinematic parameter. Significant parameters are shaded black ($p < .001$ two-sided shuffle test, n = 1000 bootstraps; error bars indicate 95% CI). **f**) Left half: Distribution of dLight waveforms across different velocity change bins. Syllable transitions were binned by the change in velocity from one syllable to the next. The peak magnitudes of dLight waveforms within each “velocity change” bin were then binned from lowest to highest; these binned dLight waveforms reveal the diversity of dLight transients associated with each behavioural transition type. Left: averaged velocity traces for each velocity change/dLight peak

bin pair. Right: averaged dLight traces for each velocity change/dLight peak bin pair. Right half: Same as left half, but transitions were binned by their associated jerk, and waveform distributions are plotted as described above. Here inter-syllable jerk is used as a surrogate for the biomechanical difficulty mice are likely to experience as they transition across syllables. **g**) Syllable-associated dopamine peaks do not contain information about position in the open field. For each syllable, peak dLight and velocity were binned into ten equally spaced bins, and the animal's 2D centroid position in the arena was binned into four equally spaced bins. Then, mutual information was computed between the dopamine and the position bin. Shown are 2D histograms of mouse position for the highest and lowest bin for dLight peaks (left) and velocity (right) for an example syllable. **h**) Per-syllable mutual information between dLight per-syllable average peaks and position in the open field ($p = .107$, $n = 57$ syllables, one-sided test). The p-value was computed by comparing the average mutual information across all syllables against the mutual information computed on shuffled data. **i**) Specific syllable transitions do not contain information about the likelihood of a dopamine transient ($p = .165$, $n = 14$ mice, one-sided test). Here, we estimated the average likelihood of syllable-associated dLight peak crossing the 95th percentile for all syllable transitions. These likelihoods were used to build a 2D matrix, where cell *i, j* was the likelihood of a transient for the transition from syllable *i* to syllable *j*. Finally, we computed the mutual information of this matrix per-mouse, and estimated p-values by comparing with the mutual information computed on shuffled data.

Extended Data Fig. 6 Dopamine predicts future syllable choices, and behaviour predicts prior dopamine dynamics.

a) Correlation matrix between dLight associated with a given syllable, entropy (which summarizes the variability of the subsequent syllable choice), syllable counts (for the syllable associated with dLight), and the dLight associated with the next syllable. Here, each feature was averaged per syllable/mouse pair, and the Pearson correlation was computed between feature averages ($n = 760$ syllable/mouse pairs). Syllable-associated dLight, entropy and syllable counts are all substantially correlated with each other, as described in the manuscript. Note that entropy (here defined as outbound entropy, the degree to which the subsequent syllable choice is predictable or

variable) does not correlate with the amount of dLight on the subsequent syllable. This observation means that the amount of dopamine associated with a given syllable does not reflect whether that specific syllable was a more or less variable choice, given the preceding syllable; this contrasts with the correlation between syllable-associated dLight and outbound entropy, which demonstrates that the amount of dopamine associated with a given syllable predicts whether the next syllable choice will be deterministic or variable. **b**) Left: schematic for an encoding model which uses future behaviour to predict average syllable-associated dLight in the past (n = 760 syllable/mouse pairs, see Methods). Middle: plot of model predictions against actual dLight peak values on held-out data (5-fold cross-validation repeated 50 times). This model combines each feature at its best lag, lag = 10 syllables for velocity, 100 syllables for counts, and 10 syllables for entropy. Each point is a syllable/mouse pair, and the color of each point represents a kernel density estimate. Regression line is shown in blue. Right: the correlation between predicted syllable-associated dLight values and actual dLight values compared to n = 1000 shuffles (average Pearson correlation of held-out mouse/syllable pairs $r = 0.46$, $p < .001$; p-values for correlations throughout this figure were estimated by comparing observed correlation to Pearson correlation from shuffled data via a one-sided test). Performance using kinematic parameters only, $r = 0.39$, counts and entropy only $r = 0.22$, both models $p < .001$ one-sided shuffle test. To evaluate model performance using feature subsets, we refit the model from scratch for each group of features using cross validation. **c**) Median beta coefficients of the encoding model shown in Extended Data Fig. 6b at increasing bin sizes. Shaded region indicates 95% confidence intervals for each behavioural variable across Markov-chain Monte Carlo samples. **d**) Schematic of a linear encoding model predicting instantaneous dLight fluorescence from future behaviour. In this model each behavioural variable is convolved with a learned kernel, with the result of each convolution summed to produce a predicted dLight trace (see Methods). **e**) Top: correlation between model predictions and true dLight fluorescence values (median correlation over all held-out experiments using all features $r = 0.28$, in black is model performance with experiment-permuted dLight traces, $p < .001$ shuffle test, n = 211 experiments). Bottom: model performance quantified as held-out correlation (2-fold cross-validation, Pearson r) shown using all behavioural variables (“all”), variables related to behavioural

structure (syllable counts or transition entropy, “syllable only”), or kinematic parameters (velocity, angular velocity, height velocity, or acceleration, “kinematic only”). Held-out correlation was evaluated for each experiment ($n = 211$). To evaluate model performance using feature subsets, we refit the model for each group of features (median r over held-out experiments for kinematic parameters 0.23; syllable-related measures 0.16; all correlations $p < .001$, one-sided shuffle test). **f**) Representative kernels learned by the fitting procedure (with cross-validation, see Methods) for each behavioural variable. Left: kernels for all behavioural variables with the same scaling. Right: kernels y-axes are re-scaled according to the scalebar shown on the left to visualize temporal dynamics of each kernel. Error bars indicate 99% bootstrap confidence interval. **g**) Model prediction of instantaneous dLight fluorescence for two example held-out experiments. Green indicates observed dLight fluctuations over time, orange indicates model-predicted fluctuations.

Extended Data Fig. 7 DMS dopamine does not correlate with syllable usage or entropy.

a) Average dopamine transient waveform for DMS and DLS. The average dopamine transient was computed for each DMS ($n = 8$) and DLS ($n = 14$) mouse, and then the per-mouse means were averaged to form the grand average. Shaded region indicates the 95% bootstrap confidence interval. As in Extended Data Fig. 1h, a transient is defined as when the dLight trace crosses 1 standard deviation above the mean. Note that DMS and DLS data are z-scored independently of each other. **b**) Summary statistics of DLS and DMS dopamine transients. Values were averaged per-mouse, thus leaving $n = 8$ for DMS and $n = 14$ for DLS. Box plots summarize per-mouse averages of each statistic. * Indicates $p < .05$, ** indicates $p < .01$, two-sided Mann-Whitney U test. Area under the curve, $p = .052$, $U = 27$, $f=.24$.; time to peak, $p = .0001$, $U = 4$, $f = .036$; Full-width at half-maximum $p = .017$, $U = 16$, $f = .14$; transient rate $p = .017$, $U = 97$, $f = .87$. **c**) As in Fig. 1f, average dorsomedial striatum (DMS) dLight fluorescence aligned to all syllable transitions (green) or movement initiations (orange) and z-scored to shuffle ($n = 100$ shuffles, see Methods; DLS shown for reference in blue). Shaded regions represent bootstrap SEM. **d**) Peak dLight values from per-mouse average waveforms aligned to either syllable transitions or movement

initiations (n = 14 DLS mice, n = 8 DMS mice). *** Indicates p < .01, two-sided Mann-Whitney U test (movement initiations, p = 2.5e-5, U = 110, f = .98; syllables, p = 4.4e-5, U = 109, f = .97). **e**) Encoding model performance for dLight peak values (see Extended Data Fig. 6b, Methods) recorded in DLS (left) and in DMS (right) using different feature subsets. Each point is the average per-mouse heldout performance (n = 14 DLS mice, n = 8 DMS mice). ** Indicates p < .01, two-sided Mann-Whitney U test. Comparison of syllable feature subsets between DLS and DMS: entropy and syllable counts p = .002, U = 102, f = 0.91; velocity p = 0.87, U = 59, f = 0.53; all p = 0.25, U = 79, f = 0.71. **f**) Correlation for all kinematic parameters for DLS (n = 14 mice) and DMS (n = 8 mice) photometry mice. Here we computed the correlation between dLight and kinematic parameters and a variety of bin sizes as in Fig. 1e. The maximum correlation across all bin sizes per mouse is shown. *, p < .05 and **, p < .01, two-sided Mann-Whitney U test. Angular velocity comparison for dLight transient rate, p = 0.006, U = 9, f = 0.08; velocity comparison for dLight transient rate, p = .03, U = 15, f = 0.13; velocity comparison for dLight average, p = 0.026, U = 14, f = 0.125.

Extended Data Fig. 8 Closed-loop Motion Sequencing.

a) Schematic of the closed-loop MoSeq pipeline. A deep neural network (a convolutional neural net (CNN)), structured as a denoising autoencoder, was used to remove small artifacts related to photometric or optogenetic fiber optics. The network takes a depth image of a mouse to 1) remove artifacts like cables, 2) eliminate rotation or translation jitter, and 3) resize the mouse to a standard size for closed loop MoSeq. This network was trained to minimize the reconstruction loss of images of mice after those images were corrupted through artificial rotation, rescaling or imposition of noise (see Methods). After being passed through this denoising network, depth frames were dimensionally reduced by applying principal components analysis, where principal components were estimated using a size-and-age-matched “clean” dataset. Finally, the principal component scores were modeled using an autoregressive hidden Markov model (AR-HMM). For offline syllable detection, discrete latent states (i.e. behavioural syllables) were estimated using the Viterbi algorithm. For online syllable detection, the probabilities of discrete latent states were estimated using a

forward pass estimated on a rolling basis (see Methods). **b**) Average pixel-wise or syllable label corruption plotted as a function of mouse depth image distortion via either a change in size or a change in rotation. Top: the impact of image corruption on syllable labels either with (orange) or without (blue) applying the CNN. Mouse depth images were corrupted through applying either a zoom factor (left, >1 indicates enlargement and <1 shrinking) or a rotation (right, in degrees), and syllable labels were compared between the corrupted and uncorrupted image using the Viterbi algorithm. Here, depth videos from a size-and-age-matched dataset were fed to the CNN after image scaling or rotation. This analysis reveals that the CNN effectively mitigates the effects of scale and rotation on the depth image. Bottom: the impact of image corruption measuring using the mean-squared-error (MSE) between the original depth image and the corrupted depth image. **c**) Similar to b, but measuring the robustness of depth images and syllable labels to jittering the mouse's position in X or Y (in units of pixels). Top: impact of position jitter on syllable labels without (left) or with (right) applying the CNN. Bottom: impact of position jitter measured using the MSE between the original depth image and the corrupted depth image without (left) or with (right) applying the CNN. **d**) Top: histogram of round-trip latency between receiving a depth frame and running all computations associated with the closed-loop pipeline (CNN, image processing, and AR-HMM likelihood estimation). Red line indicates the median syllable duration. Bottom: prediction time relative to the onset of the six syllables targeted for optogenetic reinforcement in this work. **e**) Degree to which the online system for syllable classification used during opto-DA stimulation confused the targeted syllable with other syllables. Shown is the row-normalized confusion matrix comparing online syllable calls (from actual experiments) against offline classification using traditional MoSeq (as in the remainder of the paper). The last column is the sum total of false alarms across all syllables that were not targeted for closed-loop reinforcement. **f**) Opto-DA learning is minimally impacted by false positives. Here, we show per-mouse average opto-DA learning for the targeted syllable (leftmost point of each plot), along with per-mouse averages of the 10 off-target syllables with the most false positives, ranked from highest to lowest (1 to 10). Off-target learning was smoothed with a 3-point rolling average. Results from the first stimulation experiment are shown on top, and results from the second experiment are shown on the bottom.

Extended Data Fig. 9 Closed-loop reinforcement of targeted syllables.

a) Cartoons depicting the mouse pose dynamics expressed during the six syllables targeted for optogenetic reinforcement. **b**) Per-mouse average usage plot depicting the top 40 most used syllables identified by closed-loop MoSeq (used >1% of the time), rank-ordered by baseline usage, with target syllables are outlined and highlighted. Syllable usages were computed in counts for each mouse-experiment pair, and then averaged across these pairs for each syllable (n = 32 mice total, n = 20 opto-DA mice and n = 12 controls). Target syllables are labeled in red. Error bars represent 95% bootstrapped CI. **c**) Relationship between baseline syllable usage and syllable expression duration in no-opsin controls. Each point is a syllable, whose durations and usage counts were averaged across mouse-experiment pairs, and subsequently normalized across pairs (n = 40 syllables, Spearman r = −0.08, p = 0.61). Target syllables are labeled in red. Error bars represent SEM. **d**) Probability distributions for the duration of each target syllable across all behavioural experiments and mice. Mean, median, and mode values (in seconds) are reported. **e**) Circular state map computed for the full repertoire of behaviours the closed-loop system was able to faithfully detect. Each node is a syllable, and each line represents the transition from one syllable to the next (whose width specifies the observed likelihood of each transition). Each syllable targeted for optogenetic reinforcement is shown in red; each such node is associated with a different set of sequences in which it participates. **f**) Probability distributions describing the relative timing of optogenetic stimulation offset and the offset of the syllable instance for all target syllables. Note that optogenetic stimulation across targets rarely extends into the subsequent syllable. **g**) Cumulative target syllable counts over time. Lines are averaged over the six target syllables for each mouse. Dark green indicates “learners” that used the targeted syllables significantly above controls (n = 9/20, learners are defined as mice whose average cumulative change in counts across all syllables exceeds all control mice, see Methods). **h**) Timecourse of target syllable use during the first thirteen minutes of opto-DA. Depicted here is the usage of the target syllable (in counts) above baseline in a 30-second long non-overlapping bins. Mice quickly learn the contingency between expressing the targeted

syllable and opto-DA, and then perform the target syllable at a near-constant rate above baseline.

Extended Data Fig. 10 Reinforcement of the target syllable is spatiotemporally precise.

a) Top: schematic describing the hypothesis that optogenetically-evoked DA release influences syllable counts of temporally-adjacent non-targeted syllables. Bottom: weighted average of syllable counts over baseline for non-targeted syllables for the first (left) and second (right) stimulation experiments in learner mice ($n = 9$). Green and gray shading indicates 95% bootstrap CI for weighted average and time-shuffled data, respectively ($n = 1000$ shuffles, $p > 0.05$ for all comparisons, two-sided Mann-Whitney U test). **b**) Top: schematic describing the hypothesis that opto-DA reinforces similar-velocity syllables to the target. Bottom: average syllable counts over baseline for similar-velocity syllables for stimulation experiment 1 (left) and 2 (right) in learner mice. Green and gray shading indicates 95% bootstrap CI for weighted average and time-shuffled data, respectively ($n = 1000$ shuffles, $p > 0.05$ for all comparisons, two-sided Mann-Whitney U test). **c**) Relative usage change (in syllable counts) of syllables of varying behavioural similarity to the target syllable, with syllables grouped into 10 bins given their relative similarity to the target. Shown are per-mouse-and-bin medians. Top: learner mice. Bottom: no-opsin controls. ** Indicates a significant difference between opto-DA learners ($n = 9$) and control mice ($n = 12$) ($p = 0.006$, $U = 103$, $f = 0.95$ two-sided Mann-Whitney U test between median change in counts per learner mouse), all other comparisons $p > 0.05$. **d**) Left: schematic of velocity modulation experiment (see Methods). Right: mouse/experiment averages of the targeted syllable's velocity binned by stimulation number (for velocity, $p = 0.013$, $u = 167$, $f = 0.77$; $n = 18$ up experiments and $n = 12$ down experiments, two-sided Mann-Whitney U test). Error bars indicate bootstrap SEM. **e**) Per-mouse and per-target average target syllable duration, comparing learner mice to controls. Shown is the average duration on stim trials relative to catch trials (stim – catch); no statistically significant differences in duration distributions were identified ($p = .98$ for both sessions; session one $U = 1995$ and $f = .52$; session two $U = 1804$ and $f = .46$; two-sided Mann-Whitney U test, $n = 144$ mouse/target control pairs, $n = 107$ mouse/target learner pairs). **f**) Kinematic

parameters associated with each target syllable were not altered as a result of opto-DA. Top: a linear classifier (linear discriminant analysis) was trained to use syllable-associated pose dynamics (measured using the mean and variance of the 10 principal components derived from the mouse depth data, see Methods) to predict the identity of the 6 target syllables; $p < .001$ established via a one-sided shuffle test. Bottom: linear classifiers trained on syllable-associated pose dynamics were unable to distinguish between stimulated and catch trials of single syllables in learner mice. Blue shows classifier performance on shuffled data, and red shows classifier accuracy over repeated cross validation splits; $p = 0.069$ established via a one-sided shuffle test. **g**) Stimulation of target syllables did not result in fractionated syllables or lowered detection confidence. Top: distribution showing entropy of cross-likelihoods for syllable detection for each frame, averaged across each experiment. Cross-likelihoods are a quantitative measure of confidence in assigning a given frame of behavioural data to a particular syllable. Distributions show density of average entropy of cross-likelihoods for baseline vs. stimulation experiments; these distributions show no evidence of changes in model confidence on experiments where syllables were targeted with optogenetic stimulation, consistent with opto-DA not substantially changing the kinematics associated with any given syllable in mice that learned. Bottom: distributions show probability density across baseline vs. stimulation experiments of entropy across maximum likelihoods of every syllable. No significant differences were found between stimulation and baseline distributions (all comparisons $p > .05$, two-sided 2-sample Kolmogorov-Smirnov test). **h**) Spatial histogram of frame occupancy of the centroid of the animal across stimulation and baseline experiments. Opto-DA mice (DAT-IRES-Cre::Ai32) on the left, no-opsin controls on the right. **i**) Left: Jensen-Shannon Divergence (JSD) of centroid location probability distributions across mice based on locations during stimulation trials (target performance) on stimulation day and simulated stimulation trials on baseline days ($n = 192$ mouse/target syllable pairs, $p = 0.44$, $U = 4262$, $f = 0.49$, two-sided Mann-Whitney U test across opto-DA mice and no-opsin controls). Right: JSD of centroid location distributions computed over experiment-wide centroid locations for each mouse ($n = 32$ mice, $p = 0.41$, $U = 114$, $f = 0.48$, two-sided Mann-Whitney U test). **j**) Distribution of kinematic parameters averaged per-mouse and per-target for the target syllable on both baseline and stimulation

experiments. Left: difference between stimulation and catch trials for the targeted syllable on stimulation day. Right: magnitude of kinematic parameters for all trials across baseline and stimulation experiments. No significant differences were observed between learners and controls ($p > .05$, two-sided Mann-Whitney U test). **k**) Same as right half of Extended Data Fig. 10j (for velocity and acceleration), but for all non-target syllables. No significant differences were observed between learners and controls ($p > .05$, two-sided Mann-Whitney U test). **l**) Average dLight waveform aligned to the onset of 3-second pulsed stimulation (as elicited by ChrimsonR stimulation). Gray line indicates circular shuffle. Shaded error bars indicate 95% CI. Shaded red region indicates the duration of ChrimsonR stimulation.

Extended Data Table 1 List of reagents and resources

Supplementary information

Reporting Summary

Peer Review File

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Markowitz, J.E., Gillis, W.F., Jay, M. *et al.* Spontaneous behaviour is structured by reinforcement without explicit reward. *Nature* **614**, 108–117 (2023). https://doi.org/10.1038/s41586-022-05611-2

- Received: 20 April 2022
- Accepted: 30 November 2022
- Published: 18 January 2023
- Issue Date: 02 February 2023
- DOI: https://doi.org/10.1038/s41586-022-05611-2

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article is cited by

- **Spontaneous behaviour is shaped by dopamine in two ways**
  - Dorgham Khatib

- Genela Morris

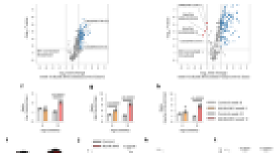
Nature (2023)

Spontaneous behaviour is shaped by dopamine in two ways

- Dorgham Khatib
- Genela Morris

Nature News & Views 18 Jan 2023

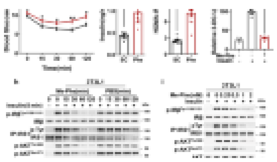
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05611-2



[Sphingolipid subtypes differentially control proinsulin processing and systemic glucose homeostasis]

21 December 2022

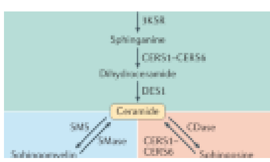
Kerstin Griess, Michael Rieck, … Bengt-Frederik Belgardt



[Phenylalanine impairs insulin signaling and inhibits glucose uptake through modification of IRβ]

25 July 2022

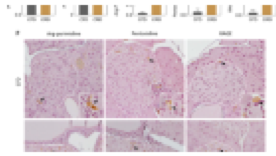
Qian Zhou, Wan-Wan Sun, … Shi-Min Zhao



[Ceramides and other sphingolipids as drivers of cardiovascular disease]

26 March 2021

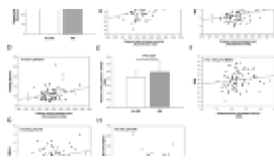
Ran Hee Choi, Sean M. Tatum, … William L. Holland



[Endogenous advanced glycation end products in pancreatic islets after short-term carbohydrate intervention in obese, diabetes-prone mice]

11 March 2019

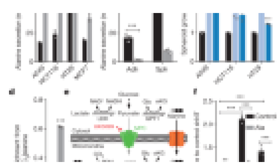
Richard Kehm, Jana Rückriemen, … Annika Höhn



[Inhibition of Serine Protease Activity Protects Against High Fat Diet-Induced Inflammation and Insulin Resistance]

03 February 2020

Chin-Sung Kuo, Jia-Shiong Chen, … Shing-Jong Lin



[Serine restriction alters sphingolipid diversity to constrain tumour growth]

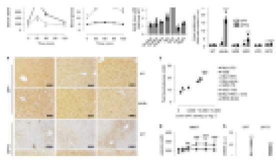
12 August 2020

Thangaselvam Muthusamy, Thekla Cordes, … Christian M. Metallo

[Role of carnitine and its derivatives in the development and management of type 2 diabetes]

07 March 2018

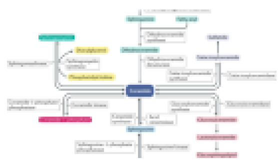
Judit Bene, Kinga Hadzsiev & Bela Melegh



[Liver alanine catabolism promotes skeletal muscle atrophy and hyperglycaemia in type 2 diabetes]

18 March 2021

Jürgen G. Okun, Patricia M. Rusu, … Adam J. Rose



[The role of ceramides in metabolic disorders: when size and localization matters]

14 February 2020

Sarah M. Turpin-Nolan & Jens C. Brüning

- Article
- Open Access
- Published: 25 January 2023

Insulin-regulated serine and lipid metabolism drive peripheral neuropathy

- Michal K. Handzlik,
- Jivani M. Gengatharan,
- Katie E. Frizzi,
- Grace H. McGregor,
- Cameron Martino,
- Gibraan Rahman,
- Antonio Gonzalez,
- Ana M. Moreno,
- Courtney R. Green,
- Lucie S. Guernsey,
- Terry Lin,
- Patrick Tseng,
- Yoichiro Ideguchi,
- Regis J. Fallon,
- Amandine Chaix,
- Satchidananda Panda,
- Prashant Mali,
- Martina Wallace,
- Rob Knight,
- Marin L. Gantner,
- Nigel A. Calcutt &
- …
- Christian M. Metallo

Show authors

Nature **volume 614**, pages 118–124 (2023)

- 21k Accesses

- 235 Altmetric

- Metrics details

Abstract

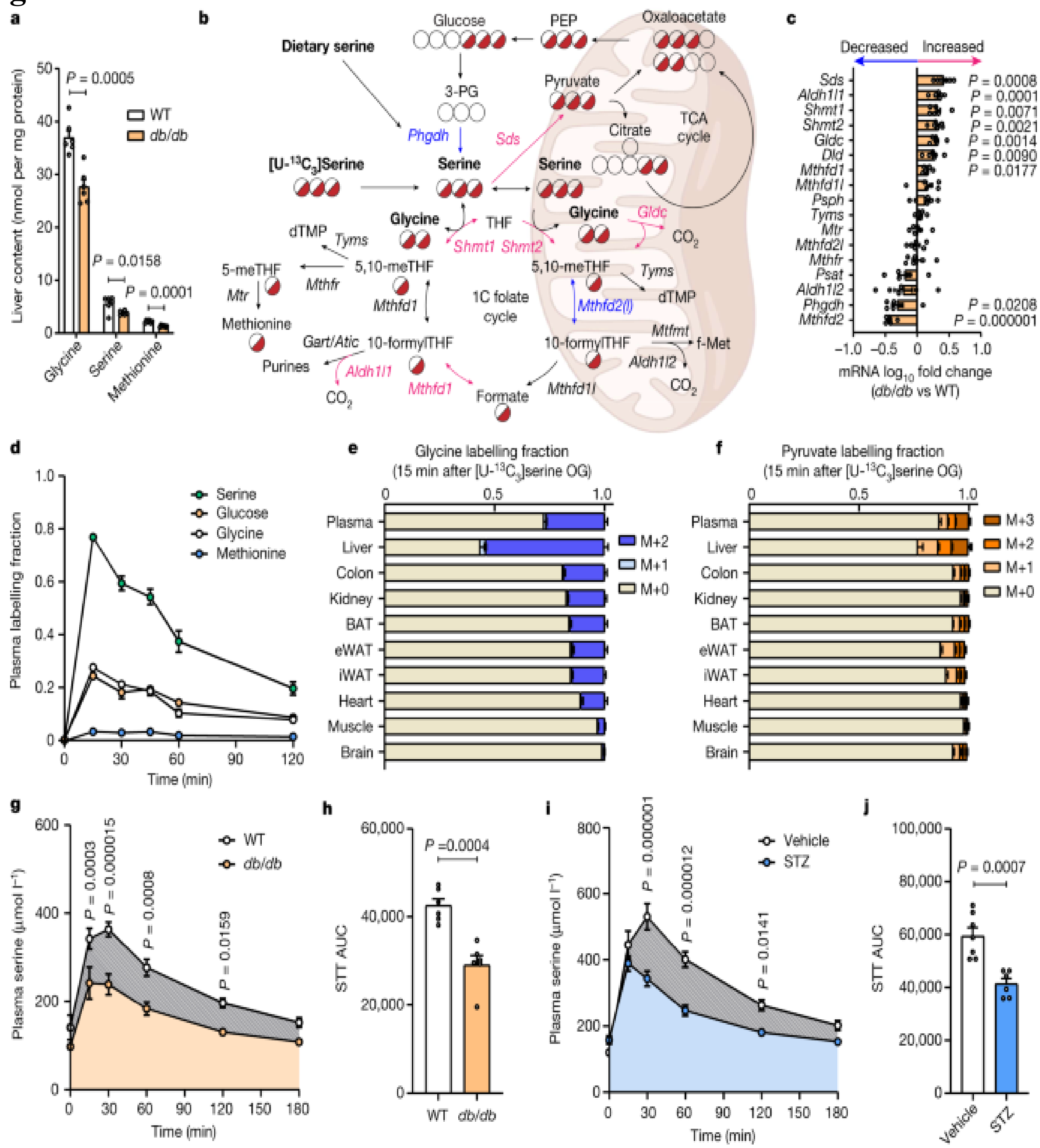
Diabetes represents a spectrum of disease in which metabolic dysfunction damages multiple organ systems including liver, kidneys and peripheral nerves[1,2]. Although the onset and progression of these co-morbidities are linked with insulin resistance, hyperglycaemia and dyslipidaemia[3,4,5,6,7], aberrant non-essential amino acid (NEAA) metabolism also contributes to the pathogenesis of diabetes[8,9,10]. Serine and glycine are closely related NEAAs whose levels are consistently reduced in patients with metabolic syndrome[10,11,12,13,14], but the mechanistic drivers and downstream consequences of this metabotype remain unclear. Low systemic serine and glycine are also emerging as a hallmark of macular and peripheral nerve disorders, correlating with impaired visual acuity and peripheral neuropathy[15,16]. Here we demonstrate that aberrant serine homeostasis drives serine and glycine deficiencies in diabetic mice, which can be diagnosed with a serine tolerance test that quantifies serine uptake and disposal. Mimicking these metabolic alterations in young mice by dietary serine or glycine restriction together with high fat intake markedly accelerates the onset of small fibre neuropathy while reducing adiposity. Normalization of serine by dietary supplementation and mitigation of dyslipidaemia with myriocin both alleviate neuropathy in diabetic mice, linking serine-associated peripheral neuropathy to sphingolipid metabolism. These findings identify systemic serine deficiency and dyslipidaemia as novel risk factors for peripheral neuropathy that may be exploited therapeutically.

Main

To explore how obesity and diabetes influence serine, glycine and one-carbon (SGOC) metabolism, we quantified serine, glycine and methionine across tissues in an established mouse model of morbid obesity, insulin resistance and hyperglycaemia (leptin receptor-deficient *db/db* mice on a black Kaliss background (BKS-*db/db*)) and compared results with age-matched wild-type C57BL/6J controls. The *db/db* mice showed reductions of around 30% in hepatic and renal serine levels relative to wild-type mice (Fig. 1a and Extended Data Fig. 1a), and the more abundant glycine pools were reduced by 30–50% in liver, kidney, inguinal white adipose tissue (iWAT) and plasma (Fig. 1a and Extended Data Fig. 1a–c). Methionine is

linked to serine through one-carbon metabolism and was also reduced in liver, iWAT, and plasma (Fig. 1a and Extended Data Fig. 1b,c), suggesting that diabetes decreases serine and glycine levels in tissues that are important for glucose and lipid homeostasis.

Fig. 1: Sources and sinks of altered serine metabolism in diabetes.



a, Levels of glycine, serine and methionine in the liver of wild-type and BKS-*db/db* mice after a 6-h fast (*n* = 6 per group). **b**, Schematic of serine

and glycine biosynthetic and catabolic pathways. Upregulated hepatic genes in BKS-*db/db* mice are in purple, and downregulated are genes are in blue. 10-formylTHF, 10-formyltetrahydrofolate; 3-PG, 3-phosphoglycerate; 5,10-meTHF, 5,10-methylenetetrahydrofolate; dTMP, deoxythymidine monophosphate; f-Met, *N*-formylmethionine; PEP, phosphoenol pyruvate; TCA, tricarboxylic acid; THF, tetrahydrofolate. **c**, mRNA expression of liver enzyme genes regulating SGOC metabolism in wild-type and BKS-*db/db* mice ($n = 6$ per group). **d**, Plasma serine, glucose, glycine and methionine-labelling fraction ($1 - M_0$) in wild-type mice administered [U-$^{13}C_3$]serine via oral gavage after an overnight fast ($n = 4$ per time point). **e**, Tissue glycine labelling fraction in wild-type mice 15 min after [U-$^{13}C_3$]serine administration via oral gavage ($n = 4$ per tissue) following an overnight fast. **f**, Tissue pyruvate labelling fraction in wild-type mice 15 min after [U-$^{13}C_3$]serine administration via oral gavage ($n = 4$ per tissue) after an overnight fast. **g**, Combined OGTT and STT in wild-type and BKS-*db/db* mice ($n = 6$ per group) after an overnight fast. **h**, STT AUC in wild-type and BKS-*db/db* mice ($n = 6$ per group). **i**, Combined OGTT and STT in vehicle- ($n = 7$) and STZ-treated ($n = 6$) C57BL/6J mice after an overnight fast. **j**, STT AUC in vehicle- ($n = 7$) and STZ-treated ($n = 6$) C57BL/6J mice. Data are mean ± s.e.m., and were analysed using two-sided independent *t*-test (**a**,**c**,**h**,**j**) and two-way ANOVA with Fisher's least significant difference post hoc test (**g**,**i**). The schematic in Fig. 1b was prepared in BioRender.

Mammals obtain serine from the diet, de novo synthesis from glucose and via glycine and one-carbon metabolism, with the liver and kidney serving as major sites for postprandial NEAA metabolism (Fig. 1b). To better understand the mechanistic basis of reduced hepatic serine and glycine in *db/db* mice, we quantified the expression of genes associated with SGOC metabolism after a 6-h fast (Fig. 1c). Genes encoding key enzymes responsible for serine and one-carbon unit catabolism or disposal were significantly upregulated in *db/db* mice, including serine dehydratase (*Sds*), serine hydroxymethyltransferase 1 (*Shmt1*), *Shmt2* and 10-formyltetrahydrofolate dehydrogenase (*Aldh1l1*). The expression of two genes encoding components of the glycine cleavage system—glycine decarboxylase (*Gldc*) and dihydrolipoamide dehydrogenase (*Dld*)—was increased in *db/db* liver, whereas the expression of 3-phosphoglycerate dehydrogenase (*Phgdh*) and methylenetetrahydrofolate dehydrogenase 2

(*Mthfd2*) were both significantly reduced, indicating that de novo serine synthesis may also be limited in diabetic mice. These results are consistent with genome-scale metabolic modelling of transcription data from human diabetic liver[17]. The kidney is another centre for SGOC metabolism[18] and renal expression of *Shmt1* and several genes encoding enzymes associated with one-carbon metabolism were increased in *db/db* mice (Extended Data Fig. 1d).

Circadian and postprandial variations in amino acids and glucose make the diagnosis of metabolic defects challenging. We therefore hypothesized that a 'serine tolerance test' (STT) could better assess SGOC metabolism in mice and identify those with elevated serine disposal. Applying similar dosages to those used in human clinical trials (ClinicalTrials.gov: NCT03062449) (400 mg kg^{-1}), we orally administered serine to overnight-fasted wild-type mice and quantified plasma serine pharmacokinetics to gauge the dynamics of serine clearance, with levels returning to baseline in approximately 2 h (Extended Data Fig. 1e). To identify the principal pathways involved in serine disposal, we orally administered [U-$^{13}C_3$]serine to overnight-fasted wild-type mice and quantified the enrichment of downstream metabolites. We observed that glucose was labelled to a similar extent as glycine throughout the test (Fig. 1d), and [U-$^{13}C_3$]serine-derived carbon was significantly incorporated into hepatic glycine, pyruvate and citrate pools (Fig. 1e,f and Extended Data Fig. 1f,g), demonstrating that hepatic gluconeogenesis is a major pathway for serine disposal in some contexts and linking its regulation to insulin and glucagon, which are both elevated in *db/db* mice (Extended Data Fig. 2a–c). To test whether insulin resistance affects serine absorption and disposal, we delivered both glucose (2 g kg^{-1}) and serine (400 mg kg^{-1}) to overnight-fasted wild-type and *db/db* mice. We observed a significant reduction in STT area under the curve (AUC) in *db/db* mice (Fig. 1g,h and Extended Data Fig. 2d), despite the higher dose administered. Notably, acute serine challenge had essentially no effect on circulating glycine concentrations (Extended Data Fig. 2e), despite evidence of their rapid interconversion (Fig. 1d,e). Conversely, there were no differences in STT AUC between wild-type and *db/db* mice when serine was administered without glucose (Extended Data Fig. 2f).

To determine whether elevated serine disposal is specific to leptin receptor-deficient *db/db* mice or is more generally attributable to impaired insulin signalling, we treated C57BL/6J mice with vehicle or streptozotocin (STZ) to induce insulin deficiency, hyperglycaemia and fat loss (Extended Data Fig. 2g–i). Plasma glycine and branched-chain amino acids (but not serine) were altered one week after STZ treatment (Extended Data Fig. 2j). Two weeks after injection, co-administration of glucose and serine revealed elevated serine disposal in STZ-diabetic mice relative to controls (Fig. 1i,j and Extended Data Fig. 2k), suggesting that insulin resistance or deficiency can both contribute to reduced circulating serine in diabetic mice.

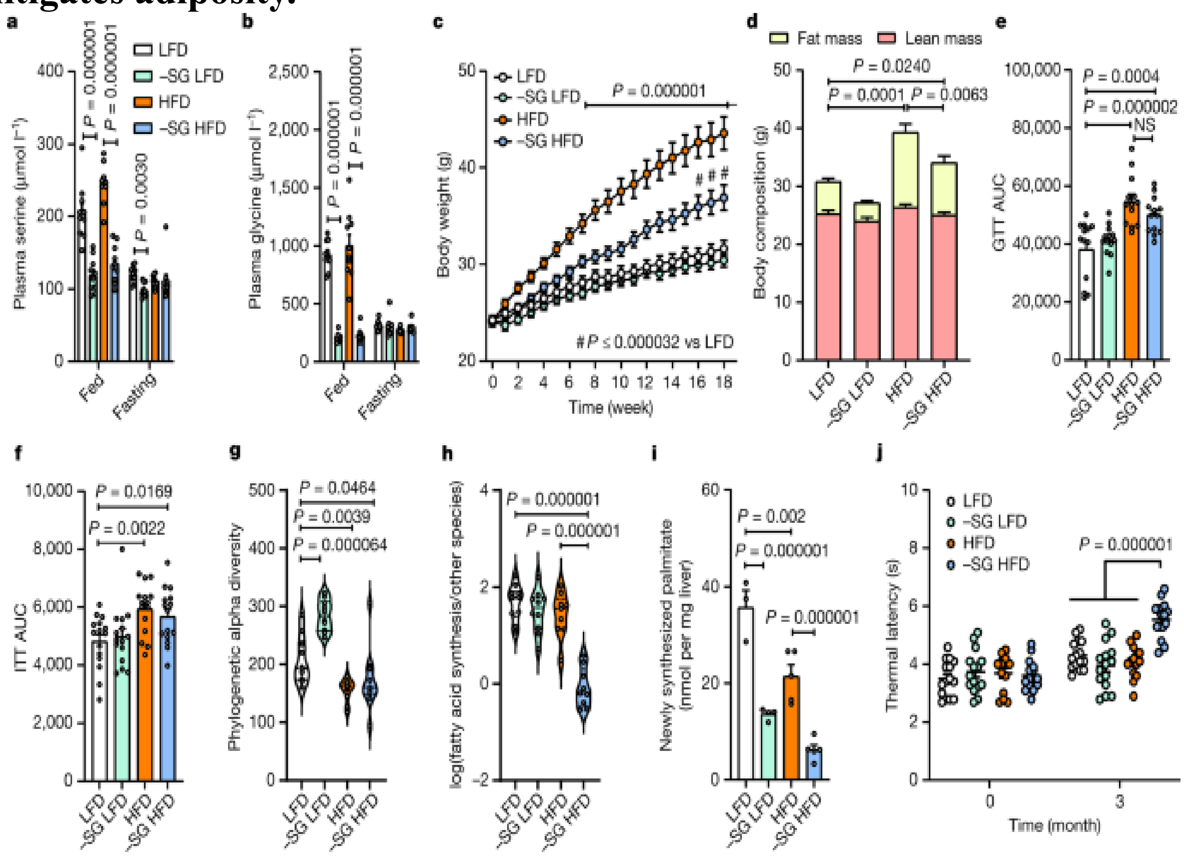
Clinical studies have implicated serine deficiency in the regulation of macular disease and peripheral neuropathy[15,16]. Given the impaired serine homeostasis in *db/db* mice, we next confirmed that aberrant serine metabolism is coincident with peripheral neuropathy in this model. Fourteen-week-old *db/db* mice exhibited thermal and tactile hypoalgesia as well as decreased motor nerve conduction velocity (MNCV) (Extended Data Fig. 2l–n). These findings suggest that aberrant serine homeostasis is associated with diabetic peripheral neuropathy.

Serine remodels lipid metabolism

Serine and glycine restriction is widely used to modulate health outcomes in mice and is relatively well-tolerated for several months[16,19,20,21,22]. To model the effect of systemic serine deficiency in the context of diet-induced obesity, we fed mice low-fat (10% kcal) (LFD) or high-fat (60% kcal from fat) (HFD) diets and compared their phenotypes to mice fed isonitrogenous diets lacking serine and glycine (−SG LFD or −SG HFD) (Supplementary Table 1). Both −SG diets effectively reduced circulating and hepatic serine and glycine contents during the fed state but not during the fasted state (Fig. 2a,b and Extended Data Fig. 3a,b). The HFD alone reduced liver glycine levels, but to a lesser extent than dietary serine and glycine withdrawal (Extended Data Fig. 3b). Other serine and glycine-derived hepatic metabolites, including glutathione, were not strongly affected by their restriction (Extended Data Fig. 3b). Of note, weight gain caused by HFD was attenuated by dietary serine and glycine restriction, whereas food, calorie and water intake, calorie absorption, and physical activity were all

unaffected (Fig. 2c and Extended Data Fig. 3c–h). Using echo magnetic resonance imaging (echoMRI), we quantified lean and fat masses across all groups and observed that dietary serine restriction significantly reduced fat mass but had no effect on lean mass relative to the HFD group (Fig. 2d). Consistent with these changes in adiposity, we observed that HFD feeding increased, whereas serine and glycine restriction decreased, epididymal white adipocyte size (Extended Data Fig. 3i).

Fig. 2: Dietary serine restriction suppresses fatty acid synthesis and mitigates adiposity.



a, Plasma serine levels in fed and overnight-fasted mice fed with LFD, −SG LFD, HFD or −SG HFD for 18 weeks (*n* = 10 per group). **b**, Plasma glycine levels in mice fed with LFD, −SG LFD, HFD or −SG HFD for 18 weeks (*n* = 10 per group). **c**, Body weight of mice fed with LFD, −SG LFD, HFD, or −SG HFD (*n* = 13 per group). **d**, Body composition in mice 18 weeks after feeding with LFD (*n* = 10), −SG LFD (*n* = 12), HFD (*n* = 13) or −SG HFD (*n* = 13). **e**, Glucose tolerance test (GTT) AUC 18 weeks after feeding with LFD, −SG LFD, HFD or −SG HFD (*n* = 13 per group). **f**, Insulin tolerance

test (ITT) AUC 18 weeks after feeding with LFD, −SG LFD, HFD or −SG HFD (*n* = 15 per group). **g**, Phylogenetic alpha diversity in mice fed with LFD, −SG LFD, HFD or −SG HFD (*n* = 10 per group). **h**, Log fraction of species from the fatty acid synthesis pathway in mice fed with LFD (*n* = 10), −SG LFD (*n* = 10), HFD (*n* = 9) or −SG HFD (*n* = 9) for 18 weeks. **i**, Hepatic de novo palmitate synthesis in mice fed with LFD (*n* = 3), −SG LFD (*n* = 4), HFD (*n* = 5) or −SG HFD (*n* = 5) for 18 weeks. **j**, Thermal sensing in mice fed with LFD (*n* = 15), −SG LFD (*n* = 15), HFD (*n* = 14) or −SG HFD (*n* = 15 per group). Data are mean ± s.e.m. and minimum and maximum (**g**,**h**), and were analysed using two-way ANOVA with Fisher's least significant difference post hoc test (**a**–**j**).

To determine how feeding with −SG HFD influences glucose homeostasis, we analysed mice using standard GTTs and ITTs. Whereas −SG HFD attenuated obesity, the mice remained glucose and insulin intolerant (Fig. 2e,f and Extended Data Fig. 3j,k), suggesting that serine and glycine restriction and the consequent reduction in adiposity do not prevent HFD-induced glucose intolerance and insulin resistance. To test whether altered systemic substrate utilization of carbohydrates in preference to lipids were driving the changes in adiposity, we placed mice that had consumed each diet for 18 weeks in metabolic cages and quantified the respiratory exchange ratio (RER). Whereas the HFD altered RER as expected, no changes were observed with dietary serine/glycine restriction (Extended Data Fig. 3l).

Next, we performed a metagenomic analysis of faecal microbiota to understand how the above diets affected microbiome composition and diversity, which correlate with and can buffer against dietary deficiencies[23]. Sustained HFD or −SG HFD feeding reduced phylogenetic alpha diversity relative to LFD-fed mice, whereas −SG LFD feeding increased alpha diversity (Fig. 2g). Conversely, PERMANOVA testing of Aitchison distances from robust principal component analysis revealed significant beta-diversity differences between −SG HFD-fed mice and other groups, highlighting the distinct effect of this low-carbohydrate, high-fat, serine and glycine-restricted diet on the faecal microbiome (Extended Data Fig. 4a). Furthermore, the log ratios of strains of microorganisms expressing complete serine biosynthesis and glycine cleavage pathways were increased and decreased, respectively, by −SG HFD (Extended Data Fig. 4b,c), and this diet reduced the log ratio of strains expressing a complete fatty acid

synthesis pathway (Fig. 2h). Key strains showing the strongest alterations are listed in Extended Data Fig. 4d. Notably, −SG LFD feeding did not modulate strains in this manner, presumably owing to the higher carbohydrate content, which could facilitate serine synthesis.

Next, to directly investigate how serine deficiency affects hepatic lipogenesis, mice fed the above diets for 18 weeks were administered heavy water (D_2O) for 18 h, and lipids were extracted for quantification of isotope enrichment, molar abundance, and synthesis[24]. Dietary serine and glycine restriction potently reduced hepatic palmitate synthesis by around 70% relative to serine-replete control diets (Fig. 2i and Extended Data Fig. 5a). Hepatic cholesterol synthesis was increased in −SG HFD compared with HFD (Extended Data Fig. 5b). Consistent with these changes, hepatic expression of ATP-citrate lyase (ACLY), acetyl-CoA carboxylase (ACC2) and stearoyl-CoA desaturase (SCD1) were strongly reduced (around 50%) by dietary serine restriction, whereas the expression of cholesterol biosynthesis enzymes was unchanged or increased (Extended Data Fig. 5c–e). Changes in AKT phosphorylation at Ser473 and Thr308 correlated with dietary fat and carbohydrate content rather than serine and glycine levels, further suggesting that serine restriction drives changes in fatty acid metabolism that are independent from insulin signalling (Extended Data Fig. 5c).

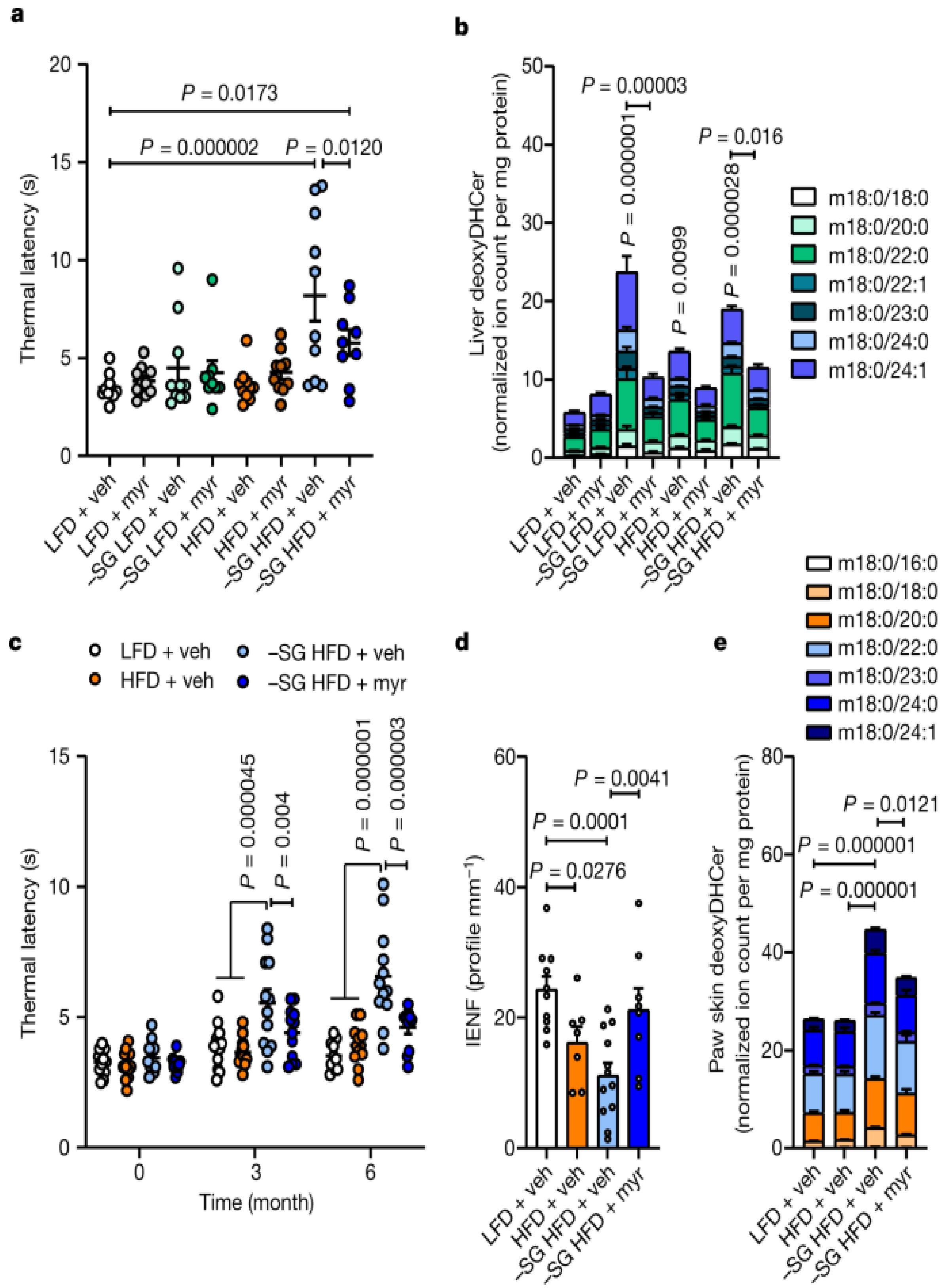
Serine restriction and HFD accelerate neuropathy

Systemic serine deficiency has recently been linked to various neurodegenerative disorders[16,25,26,27,28]. People with diabetes who have elevated serine disposal could therefore be more susceptible to neurological co-morbidities reminiscent of serine-associated peripheral neuropathy. To determine whether long-term, chronic serine deficiency is sufficient to drive peripheral neuropathy, we fed mice either control or serine- and glycine-free chow diets (19.2% of energy from fat) for up to 12 months. Temporal quantification of the thermal response to heat revealed progression to hypoalgesia after 12 months of dietary intervention (Extended Data Fig. 6a), consistent with previous observations[16]. At this time point we also detected reduced intraepidermal nerve fibre (IENF) density in paw skin (Extended Data Fig. 6b), which is also a clinical measure of small sensory fibre

degeneration[29]. Of note, mice fed −SG HFD for just three months exhibited marked thermal hypoalgesia (Fig. 2j), indicating that a combination of low systemic serine and HFD accelerate the onset of peripheral neuropathy in mice.

Serine is essential for the biosynthesis of canonical sphingolipids, which are enriched in the nervous system. When serine becomes limited, serine palmitoyltransferase (SPT) incorporates other amino acids, including alanine, to form non-canonical deoxysphingolipids[30,31]. Given the importance of canonical ceramides and 1-deoxy(dihydro)ceramides in obesity and neuropathy, respectively[15,32], we hypothesized that SPT inhibition could influence the observed obesity and neuropathy phenotypes. We therefore administered myriocin (0.3 mg kg^{-1} every other day), an inhibitor of SPT, or vehicle to mice fed the above diets for 6 months and quantified sphingolipid diversity and thermal sensing. Consistent with previous reports[32], myriocin treatment attenuated HFD-induced weight gain without affecting plasma serine and glycine levels (Extended Data Fig. 6c–f) However, myriocin also mitigated thermal hypoalgesia exhibited by mice fed −SG HFD (Fig. 3a), suggesting that a reduction in SPT activity reduced serine-associated peripheral neuropathy.

Fig. 3: Inhibition of de novo sphingolipid biosynthesis decelerates the kinetics of serine-associated peripheral neuropathy.



a
Thermal latency (s)
P = 0.0173
P = 0.000002
P = 0.0120
LFD + veh
LFD + myr
–SG LFD + veh
–SG LFD + myr
HFD + veh
HFD + myr
–SG HFD + veh
–SG HFD + myr
b
Liver deoxyDHCer (normalized ion count per mg protein)
P = 0.00003
P = 0.000001
P = 0.0099
P = 0.016
P = 0.000028
m18:0/18:0
m18:0/20:0
m18:0/22:0
m18:0/22:1
m18:0/23:0
m18:0/24:0
m18:0/24:1
c
LFD + veh
HFD + veh
–SG HFD + veh
–SG HFD + myr
Thermal latency (s)
P = 0.000045
P = 0.004
P = 0.000001
P = 0.000003
Time (month)
d
IENF (profile mm^{-1})
P = 0.0041
P = 0.0001
P = 0.0276
e
m18:0/16:0
m18:0/18:0
m18:0/20:0
m18:0/22:0
m18:0/23:0
m18:0/24:0
m18:0/24:1
Paw skin deoxyDHCer (normalized ion count per mg protein)
P = 0.0121
P = 0.000001
P = 0.000001

a, Thermal latency in mice fed with LFD plus vehicle (veh) ($n = 10$), LFD plus 0.3 mg kg^{-1} myriocin (myr) ($n = 10$), −SG LFD plus vehicle ($n = 10$), −SG LFD plus myriocin ($n = 9$), HFD plus vehicle ($n = 10$), HFD plus myriocin ($n = 10$), −SG HFD plus vehicle ($n = 10$) or −SG HFD plus myriocin ($n = 9$). **b**, Stack plot of liver deoxyDHCer in mice fed with LFD plus vehicle ($n = 10$), LFD plus myriocin ($n = 10$), −SG LFD plus vehicle ($n = 10$), −SG LFD plus myriocin ($n = 9$), HFD plus vehicle ($n = 10$), HFD plus myriocin ($n = 10$), −SG HFD plus vehicle ($n = 10$) or −SG HFD plus myriocin ($n = 9$). **c**, Thermal latency time course in mice fed with LFD plus vehicle ($n = 12$), HFD plus vehicle ($n = 12$), −SG HFD plus vehicle ($n = 12$) or −SG HFD plus myriocin ($n = 11$). **d**, IENF density in mice fed with LFD plus vehicle ($n = 10$), HFD plus vehicle ($n = 7$), −SG HFD plus vehicle ($n = 12$) or −SG HFD plus myriocin ($n = 8$). **e**, Paw skin deoxyDHCer distribution in mice fed with LFD plus vehicle ($n = 12$), HFD plus vehicle ($n = 12$), −SG HFD plus vehicle ($n = 12$) or −SG HFD plus myriocin ($n = 11$). Data are mean ± s.e.m., and were analysed using one-way ANOVA with Fisher's least significant difference post hoc test (**a**,**b**,**d**,**e**) or two-way ANOVA with Fisher's least significant difference post hoc test (**c**). Statistical analyses in **b**,**e** were performed using summed deoxyDHCer abundances.

To better understand the metabolic drivers of this peripheral neuropathy phenotype, we next quantified ceramides and deoxydihydroceramides (deoxyDHCer) in liver and sciatic nerve. Restriction of serine and glycine increased deoxyDHCer in LFD and HFD settings and reduced canonical ceramides in HFD background, whereas myriocin generally reduced abundances of sphingolipids (Fig. 3b and Extended Data Fig. 6g–i). By contrast, ceramide and 1-deoxysphingolipid levels were unaltered or did not correlate with peripheral neuropathy phenotype in sciatic nerve, potentially owing to the large lipid deposits present in myelin (Extended Data Fig. 6h,i). Notably, serine restriction in a LFD background did not induce thermal hypoalgesia after six months (Fig. 3a), indicating that 1-deoxysphingolipid accumulation alone is insufficient to drive this phenotype and consistent with recent reports of an *Sptlc1*^{C133W} mouse model[33]. Next, we measured IENF density, corneal nerve density, and tissue lipids in a separate cohort of mice fed LFD, HFD, −SG HFD or −SG HFD plus myriocin for six months. Myriocin mitigated the onset of thermal hypoalgesia in mice fed −SG HFD

(Fig. 3c), and this treatment also protected small fibre nerve density in the epidermis and cornea (Fig. 3d and Extended Data Fig. 7a,b) without affecting tactile sensing or MNCV (Extended Data Fig. 7c,d), suggesting this early-onset phenotype is specific to small sensory fibres and 1-deoxysphingolipids, in contrast to other nodes in sphingolipid metabolism[34]. Mice fed −SG HFD exhibited increased hepatic 1-deoxysphingolipids and sphingomyelin, whereas myriocin strongly reduced the levels of sphingolipids as well as triglycerides and diacylglycerides (Extended Data Fig. 7e), further highlighting the pleiotropic effects of this molecule on the lipidome[32,35]. Finally, paw skin exhibited a significant increase in deoxyDHCer that was reduced with myriocin treatment (Fig. 3e), suggesting the lipid microenvironment surrounding small fibres can influence sensory function.

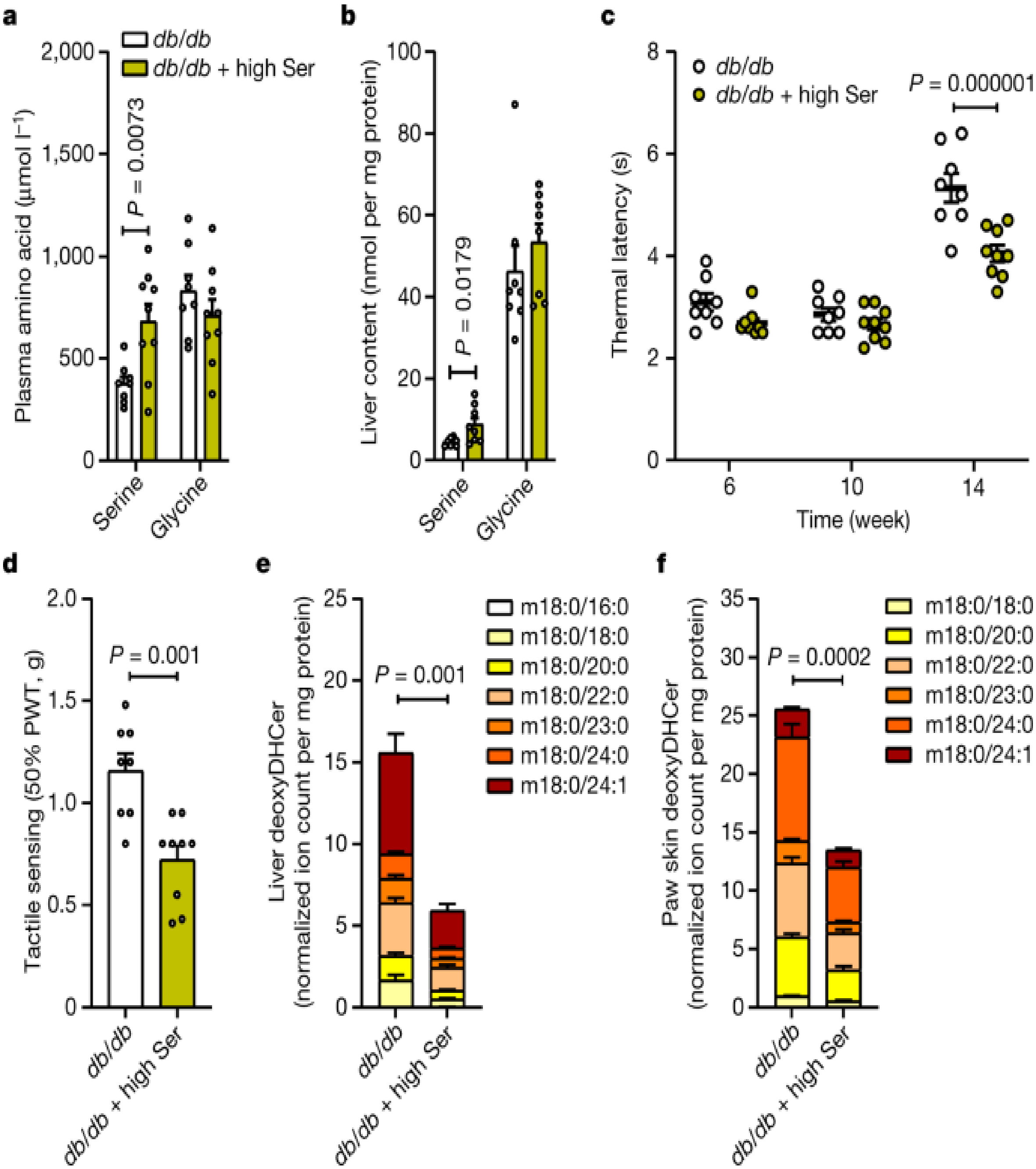
We next evaluated whether suppressing SPT activity could mitigate neuropathy in *db/db* mice, which have increased circulating deoxySA as well as increased hepatic ceramides and deoxyDHCer (Extended Data Fig. 8a,b). Consistent with results in the −SG HFD diet-induced peripheral neuropathy model, administration of myriocin at six weeks of age prevented progression to thermal hypoalgesia and restored tactile sensation in *db/db* mice (Extended Data Fig. 8c,d). IENF density was also increased in *db/db* mice dosed for eight weeks with myriocin (Extended Data Fig. 8e). Although myriocin did not affect body weight gain, hyperglycaemia or plasma serine levels (Extended Data Fig. 8f–h), it strongly reduced canonical sphingolipids in the liver but showed limited effects on paw skin 1-deoxysphingolipids and ceramides (Extended Data Fig. 8i–n). Thus, myriocin probably acts through both direct and indirect mechanisms targeting liver and other tissues, which also accounts for its toxicity[35].

Serine supplementation slows neuropathy

Considering the chronic serine deficiency exhibited by diabetic mice, we next fed *db/db* mice a 3% serine-enriched diet starting at 6 weeks of age and quantified neuropathy phenotypes. Thermal sensation was measured at 6, 10 and 14 weeks of age and tactile sensation at time of sacrifice, at which point mice exhibited elevated plasma and hepatic serine, but not glycine, levels (Fig. 4a,b). We observed no change in body weight and a slight increase in

circulating glucose levels in this cohort (Extended Data Fig. 9a,b). However, both thermal and tactile hypoalgesia were reduced in mice fed this serine-enriched diet (Fig. 4c,d). Of note, no change in canonical sphingolipids was observed across tissues, yet 1-deoxysphingolipids were robustly decreased in both liver and paw skin (Fig. 4e,f and Extended Data Fig. 9c–f). Collectively, these data suggest that supplementation of serine can slow the progression of diabetic peripheral neuropathy.

Fig. 4: Dietary serine supplementation reduces deoxysphingolipid content and slows down peripheral neuropathy.



a, Plasma amino acid levels in the fed state in BKS-*db/db* mice fed with either a control ($n = 8$) or serine-supplemented diet ($n = 9$) for 8 weeks. **b**, Liver amino acid content in the fed state in BKS-*db/db* mice fed with either a control or serine-supplemented diet for 8 weeks ($n = 8$ per group). **c**, Thermal latency time course in BKS-*db/db* mice fed with either a control ($n = 8$) or serine-supplemented diet ($n = 9$). **d**, Tactile sensing 8 weeks after feeding BKS-*db/db* mice with either a control ($n = 8$) or serine-supplemented diet ($n = 9$). **e**, Levels of deoxyDHCer in the liver in BKS-*db/db* mice fed

with either a control or serine-supplemented diet for 8 weeks ($n = 8$ per group). **f**, Levels of deoxyDHCer in the paw skin of BKS-*db/db* mice fed with either a control or serine-supplemented diet for 8 weeks ($n = 8$ per group). Data are mean ± s.e.m., and were analysed using a two-sided independent *t*-test (**a**,**b**,**d**–**f**) or two-way ANOVA with Fisher's least significant difference post hoc test (**c**). Statistical analyses in **e**,**f** were performed using summed deoxyDHCer abundances.

Discussion

Here we describe how direct or indirect induction of chronic, systemic serine deficiency alters lipid homeostasis and contributes to diabetic peripheral neuropathy. Modulating dyslipidaemia with myriocin or 1-deoxysphingolipid biosynthesis with serine supplementation both mitigate thermal and tactile hypoalgesia in obese diabetic mice. These results highlight how serine deficiency can synergize with dyslipidaemia to alter neurological phenotypes both in rare disease contexts[16,25,26] and indirectly via a widespread, chronic disease such as type 2 diabetes, where it manifests as a co-morbidity experienced by a subset of patients. Reduced circulating serine and glycine in diabetic mice may be driven by increased flux through gluconeogenesis, one-carbon metabolism, renal retention[18,36] and/or disposal as acylglycines[37], which are also influenced by dyslipidaemia[38].

A STT, analogous to an oral glucose tolerance test (OGTT), could identify patients that exhibit elevated, postprandial serine disposal and who might be particularly susceptible to sensory neuropathy. Normalizing circulating serine levels via dietary supplementation delays the onset and progression of sensory neuropathy in *db/db* mice. Indeed, supplementation of serine and B vitamins improves peripheral neuropathy in some pre-clinical models and are the focus of clinical trials for various neurodegenerative disorders[39,40,41] (ClinicalTrials.gov: NCT03062449). Our results highlight physiologically relevant molecular links between serine and glycine homeostasis, sphingolipid metabolism, and diabetic co-morbidities. The metabolic and neuropathic phenotypes of mouse models of diabetes and obesity vary across strains and genotypes[42,43], and C57BL6/J mice are particularly sensitive to the metabolic consequences of HFD owing to mutations in *Nnt*[44] and other genes. However, our data in both diabetic *db/db* mice and in C57BL/6J mice

fed −SG HFD demonstrate that serine deficiency combined with dyslipidaemia can drive peripheral neuropathy in different genetic backgrounds.

Several key questions remain, including why serine and glycine deprivation suppresses hepatic fatty acid synthesis and gene expression in the liver. In addition, diverse sphingolipid species and/or their mis-localization contribute to neuropathy[21,45,46,47,48], but their exact mechanisms of toxicity remain unclear. We have developed and validated a dietary model of serine-associated sensory neuropathy that develops a phenotype in three months, which could aid in understanding how neurotoxic dyslipidaemia can be managed. Diverse genetic changes may influence circulating serine and glycine, including common single nucleotide polymorphisms and rare coding events[25,28,49], or such deficiencies could be induced by diabetes-induced rewired hepatic metabolism. Thus, in both altering sphingolipid diversity and compromising the liver's ability to handle nutritional lipid overload, systemic serine deficiency emerges as a modifier of age- and diabetes-associated neuropathies.

Methods

Mouse experiments

All mouse experiments were approved and conducted in accordance with the Institutional Animal Care and Use Committee (IACUC) of the University of California, San Diego and the Salk Institute for Biological Studies. Mice were housed in the same room ensuring exposure to the same temperature (21 °C), humidity (ambient humidity 65%) and a 12-h light:dark cycle (06:00–18:00). In Fig. 1, 14- to 16-week-old C57BL/6J (JAX 000664) or BKS-*db/db* mice (JAX 000642), and 10- to 12-week-old vehicle- or STZ-treated C57BL/6J (JAX 000664) mice were fasted for 6 h prior to tissue collection. Animals were anaesthetized with isoflurane, decapitated, and tissues were rapidly collected using Wollenberger clamps pre-cooled to the temperature of liquid nitrogen and stored at −80 °C until analysis. For Fig. 2, 8-week-old C57BL/6J (JAX 000664) were fed with diets obtained from Envigo. Dietary composition is detailed in Supplementary Table 1. In dietary experiments, tissues were collected between 07:00–10:00 h—that is, in the

fed state unless stated otherwise. From a separate cohort of animals, plasma samples were collected 18 weeks after dietary intervention in the fed (07:00–10:00 h) and fasted (18-h overnight fast) state. In *db/db* mice experiments, tissues were collected after 6-h fasting. For Fig. 3, 8-week-old C57BL/6J (JAX 000664) were fed with diets obtained from Envigo. Tissue collection took place between 07:00–10:00 h. Animals were anaesthetized with isoflurane, decapitated, and tissues were rapidly collected using Wollenberger clamps pre-cooled to the temperature of liquid nitrogen and stored at −80 °C until analysis. For Fig. 4, 6-week-old BKS-*db/db* mice (JAX 000642) were fed with either a control or serine-supplemented diet (provided by Envigo) for a period of 8 weeks. Tissue collection took place between 07:00–10:00 h unless stated otherwise. Animals were anaesthetized with isoflurane, decapitated, and tissues rapidly collected using Wollenberger clamps pre-cooled to the temperature of liquid nitrogen and stored at −80 °C until analysis.

Serine tolerance test

Age-matched 14- to 16-week-old wild-type and BKS-*db/db*, and 10- to 12-week-old vehicle- and STZ-treated C57BL/6J (JAX 000664) mice were fasted overnight with water access provided ad libitum. For a STT, animals were weighed, and serine and/or glucose were administered via oral gavage at a dose of 400 mg kg^{-1} and 2 g kg^{-1}, respectively, with tail tip blood samples collected into EDTA-coated microvette tubes (Sarstedt) before, and 15, 30, 60, 120 and 180 min after an oral gavage. EDTA microvettes were spun at 2,000*g* at 4 °C for 5 min to obtain plasma, and samples stored at −80 °C until analysis. Blood glucose and serine concentrations were quantified using Contour Next glucometer (Bayer) and gas chromatography–mass spectrometry as described below, respectively. Plasma serine pharmacokinetics were determined for a 400 mg kg^{-1} dose using PK solver[50].

To qualify downstream fate of serine, wild-type mice were fasted overnight, weighed in the morning, and [U-$^{13}C_3$]serine administered via oral gavage at a dose of 400 mg kg^{-1}, with tissues collected, using Wollenberger clamps pre-cooled to the temperature of liquid nitrogen, before, and 15, 30, 45, 60, and 120 min after oral gavage, and samples stored at −80 °C until analysis.

Serum insulin and glucagon measurements

Commercially available kits were used to determine serum insulin (Mouse Insulin ELISA 10-1247-01, Mercodia) and glucagon (Glucagon ELISA 10-1271-01, Mercodia) following a 6-h fast in mice according to the manufacturer's instructions.

Lipogenesis D_2O experiments

C57BL/6J mice fed diets for 18 weeks were injected intraperitoneally with D_2O (in 0.9% NaCl) at a dose of 0.027 ml per g of body weight with drinking water replaced with 6% D_2O-enriched solution for a period of ~18 h. In the morning (07:00–10:00 h) tissues were rapidly collected using Wollenberger clamps pre-cooled to the temperature of liquid nitrogen and stored at −80 °C until analysis.

Plasma D_2O enrichment was determined using deuterium–acetone exchange protocol as previously described[24]. In brief, 5 µl of plasma were incubated with 4 µl of 5% acetone in acetonitrile solution and 4 µl of 10 M NaOH for 24 h. Next, 500 mg of Na_2SO_4 and 600 µl of chloroform were added, and samples vortex-mixed. After 2 min centrifugation at 3,000*g*, 80 µl was transferred in triplicate into gas chromatography–mass spectrometry (GC–MS) vials, and plasma D_2O enrichment was quantified from an external standard curve on an Agilent DB-35MS column (30 m by 0.25 mm internal diameter × 0.25 µm, Agilent J&W Scientific) installed in an Agilent 7890 A gas chromatograph (GC) interfaced with an Agilent 5975 C mass spectrometer with the following temperature program: 60 °C initial, increase by 20 °C min^{-1} to 100 °C, increase by 50 °C min^{-1} to 220 °C, and hold for 1 min.

To quantify tissue D_2O labelling, ~20 mg of frozen tissue was homogenized with 250 µl −20 °C methanol, 250 µl ice-cold saline and 500 µl −20 °C chloroform spiked with internal standards palmitate-d_{31} (Cambridge Isotope Laboratories, DLM-215-PK) and coprostanol (Sigma, 7578). After a 5 min spin at 4 °C at 21,000*g*, the chloroform fraction was collected, dried, and resuspended with 500 µl of 2% H_2SO_4 in methanol for 2 h at 50 °C. Next,

100 µl of saturated NaCl and 500 µl of hexane were added, samples vortex-mixed, and upper hexane phase collected and transferred into a GC–MS vial. Fatty acid methyl esters were analysed using a Select FAME column (100 m × 0.25 mm internal diameter) installed in an Agilent 7890 A GC interfaced with an Agilent 5975 C MS using the following temperature program: 80 °C initial, increase by 20 °C min^{-1} to 170 °C, increase by 1 °C min^{-1} to 204 °C, then 20 °C min^{-1} to 250 °C and hold for 10 min. The percent isotopologue distribution of each fatty acid and polar metabolite was determined and corrected for natural abundance using in-house algorithms adapted from a previous report[51].

GTT and ITT

For GTT and ITT, C57BL/6J mice fed the diets for 18 weeks were fasted overnight with water provided ad libitum. In the morning animals were weighed and fasting blood glucose was determined from a tail bleed using a Contour Next glucometer (Bayer). For GTT, the animals were injected intraperitoneally with a bolus of glucose at a dose of 2 g kg^{-1} of body weight, and blood glucose determined at 15, 30, 60, 120 and 180 min post-injection. For ITT, the animals were injected intraperitoneally with a bolus of insulin (100 IU ml^{-1} Humulin Insulin, Eli Lilly) at a dose of 0.5 IU kg^{-1}, and blood glucose was quantified at 15, 30, 60 and 90 min post-injection as previously described[52].

Determination of body composition and systemic metabolic rates

Lean and fat masses were determined using a EchoMRI 3-in-1 instrument (quantitative nuclear magnetic resonance (qNMR) imaging system). Comprehensive Laboratory Animal Monitoring System (CLAMS) (Oxymax, Columbus Instruments) was used to quantify systemic metabolic rates in individually housed mice during a period of 6 days. Water, food, and calorie intake were calculated from individually housed animals over a period of 6 days when subjected to CLAMS. Whole-body oxygen consumption (VO_2) and carbon dioxide (VCO_2) rates were normalized to

corresponding total body weights, and RER was calculated as ratio of VCO_2 to VO_2.

Faecal bomb calorimetry and calorie absorption

Approximately 1 g faeces was desiccated overnight and ground using a mortar and pestle. Powdered sample was reconstituted into a pellet with 300 µl ddH_2O and weighed. Pellet was placed in bomb cylinder surrounded by 2,000 ml ddH_2O (Parr 6100 Compensated Jacket Calorimeter). Heat produced by combustion was measured by change in water temperature. The calorimeter energy equivalent, W (Cal $°C^{-1}$), was calculated with standardized benzoic acid. The final energy content of each pellet was calculated as follows:

$${\rm{Gross}}\,{\rm{energy}}=\frac{({T}_{{\rm{final}}}\,\mbox{--}{T}_{{\rm{initial}}})}{{\rm{sample}}\,{\rm{weight}}}$$

Calorie absorption was calculated by subtracting gross energy (faecal calorie extraction) from calorie intake.

Microbiome analysis

DNA was extracted from 10–30 mg of stool using the MoBio PowerFecal DNA isolation kit (12830-50). Extracted DNA was quantified using a Nanodrop (ThermoFisher Scientific). The whole-genome sequencing raw data was uploaded to Qiita[53], where we followed their default processing workflow. In summary, the raw reads were adapter filtered using the auto-detect parameters in fastp version 20[54] and host (mouse) filtered using minimap2 version 2.17[55]. The resulting sequences were aligned using Bowtie 2 version 2.4.2[56] to the Web of Life (WoL) reference database[57] via the Web of Life Toolkit App (https://github.com/qiyunzhu/woltka); this step generated tables at genus, species, per genome, and per gene tables. For all analyses we used the per genome table; then for alpha diversity we removed any samples below 1,273,062 sequences per sample and for beta-diversity analysis we rarefied at the same value. Downstream analyses were performed in QIIME 2 version 2020.11[58]. To asses global microbiota

alterations, alpha diversity analysis was performed through Faith's PD[59] and beta diversity through robust principal component analysis (RPCA)[60] and resulting Aitchison distances were evaluated through permutational multivariate analysis of variance (PERMANOVA)[61].

We then designed a Bayesian hierarchical model for differential abundance incorporating diet type as a fixed effect and cage as a random effect. We model the count generating process as a negative binomial distribution to account for overdispersion. Due to the sparsity of microbiome data, we also accounted for zero-inflation by assigning each microbe a probability of being unobserved separately from the count generating process:

$${y}_{ij}=\left\{\begin{array}{cl}0,\quad & {\theta }_{j}=1\\ {\rm{Negative}}\,{\rm{binomial}}({\eta }_{ij},{\varphi }_{j}),\quad & {\theta }_{j}=0\end{array}\right.$$

$${\theta }_{j}={\rm{Bernoulli}}({\pi }_{j})$$

$$\log ({\eta }_{ij})={x}_{i}{\beta }_{j}+{z}_{i}{u}_{j}+\,\log ({{\rm{depth}}}_{i})$$

We wrote this model using the Stan probabilistic programming language[62] and fit the model using BIRDMAn (https://github.com/gibsramen/BIRDMAn). To account for compositionality, we fit this model using the first microbe in the table as an additive log ratio reference and converted log fold changes into centred log ratio coordinates after fitting. We used the following as prior distributions for the target parameters:

$$\begin{array}{c}{\pi }_{j}\sim {\rm{Beta}}(1.5,1.5)\\ {\varphi }_{j}^{-1}\sim {{\rm{Cauchy}}}_{+}(0,3)\\ {\beta }_{j}\sim {\rm{Normal}}(0,5)\\ {u}_{j}\sim {\rm{Normal}}(0,2)\end{array}$$

in which i is the sample, j is the feature, y is the microbial count, θ is the indicator for non-biological zero, η is the mean feature count, x is the covariate, β is the regression coefficients to be estimated (log-fold changes), π is the probability of non-biological zero, z is the cage identifier variable, u is the random effect of cage, and ϕ is the overdispersion parameter. In order to compare functional changes associated with strain level differential abundances a comparative genomics pathway completeness

approach was taken. First, each genome was assessed via MetaCyc[63] pathway completeness, a proportion ranging from zero to one, by mapping characterized genes to reactions and finally to pathways. Each pathway was then correlated by Spearman's rank correlation to the beta differential abundance determined from the above model. Serine biosynthesis, glycine cleavage, and fatty acid synthesis pathways were significantly correlated to betas. To validate these correlations, the log ratio of the sum of the abundance of genomes with complete pathways (completeness = 1) vs. those without (completeness < 1) were evaluated between treatment groups.

Behavioural assays

Thermal sensation

Small sensory C fibre function was quantified by behavioural responses to heat using a thermal nociception test device (UARD) as previously described[64]. In brief, the apparatus surface was warmed up to 30 °C, and animals were placed in individual testing chambers for 20–30 min prior to testing. Four separate response latency measurements were performed, and the mean of the last triplicate taken to represent response latency for each animal. All measurements were made on coded animals by an observer unaware of the treatment groups.

Tactile sensation

Animals were placed on the von Frey stand and allowed to acclimate for 20–30 min. The range of manual von Frey filaments was used: 2.44, 2.83, 3.22, 3.61, 3.84, 4.08, 4.31, 4.56, 4.74 (Kom Kare). Testing began with the 3.84 filament and the pressure applied was repeated five times. If a positive response was observed, the next lower weighted filament was used in the sequence. In the case of a negative response, the next higher weighted filament was applied. All measurements were made on coded animals by an observer unaware of the treatment groups.

Nerve conduction velocity

Conduction of a motor nerve was quantified in anaesthetized mice using EZ Anesthesia Versaflex system (Braintree Scientific, Z-7150). In brief, lightly anaesthetized mice were transferred onto a water-heated pad with anaesthesia maintained via a face mask. Two recording platinum electrodes were inserted between the animal's second, third, and fourth toes, and a grounding electrode into the skin at the neck. PowerLab stimulator delivered a 200-mV, 50-µs square-wave stimulus every 2 s. The stimulating electrode was inserted in the ankle near the Achilles tendon and subsequently into the sciatic notch at the hip, and M waves were recorded. The latency between Achilles tendon and sciatic notch was used to calculate nerve conduction velocity as described[64]. All measurements were made on coded animals by an observer unaware of the treatment groups.

Corneal confocal imaging

Quantification of corneal nerves was performed in anaesthetized mice (using EZ Anesthesia Versaflex system, Braintree Scientific, Z-7150) using Retina Tomograph 3 with Rostock Cornea Module (Heidelberg Engineering) equipped with Tomocap (Heidelberg Engineering, 0220-001) as previously described[64]. In brief, lightly anaesthetized mice were transferred onto a small animal platform with anaesthesia maintained via a face mask. Forty sequential images of uniform magnification and size were collected and those containing nerves of the sub-basal plexus identified. ImageJ software (ImageJ 1.53e Java 1.8.0_172) was used to quantify corneal nerve area within each image, with data presented as pixels/image. All measurements were made on coded animals and images by an observer unaware of the treatment groups.

Epidermal innervation

Quantification of epidermal innervation was performed in paw skin samples by immunostaining for the pan-neuronal protein PGP9.5, as described previously in detail[64]. In brief, paw skin samples were collected into 4% buffered paraformaldehyde (Thermo scientific, J19943-K2). Staining of epidermal nerves was performed using anti-PGP9.5 antibody (ProteinTech, 14730-1-AP; 1:500 dilution). Using 40× magnification of a light microscope, the number of PGP9.5-positive profiles present in the epidermis

was calculated, length of skin section calculated, and IENF profile density expressed as profiles mm^{-1}. All measurements were made on coded slides by an observer unaware of the treatment groups.

Metabolite extraction and quantification

Plasma polar metabolites were extracted from 3 µl of plasma spiked with a known amount of ^{13}C- and ^{15}N-labelled standards (Cambridge Isotope Laboratories, MSK-A2-1.2). Tissue metabolite extraction was performed as described before[21]. In brief, ~20 mg of tissue was homogenized for 2 min using Precellys beads with 500 µl −20 °C methanol, 400 µl ice-cold saline and 100 µl ice-cold water and spiked with ^{13}C/^{15}N polar metabolite standards (Cambridge Isotope Laboratories, MSK-A2-1.2), 20 pmol of sphinganine-d7 (Avanti Polar Lipids, 860658), 2 pmol of deoxysphinganine-d3 (Avanti Polar Lipids, 860474), 100 pmol of ^{13}C-dihydroceramide-d7 (Avanti Polar Lipids, 330726), 200 pmol of C_{15}-ceramide-d7 (Avanti Polar Lipids, 860681), 10 pmol of d18:1-d7 glucosylsphingosine (Avanti Polar Lipids, 860695), 100 pmol of d18:1-d7/15:0 glucosylceramide (Avanti Polar Lipids, 330729), 100 pmol of d18:1-d7/15:0 lactosylceramide (Avanti Polar Lipids, 330727), 200 pmol of sphingosine-d7 (Avanti Polar Lipids, 860657), and 200 pmol of d18:1/18:1-d9 sphingomyelin (Avanti Polar Lipids, 791649). The identification of 1-deoxydihydroceramides was confirmed via retention time matching and analysis of m18:0/24:1 deoxyDHCer (Avanti Polar Lipids, 860464) and m18:0/16:0 deoxyDHCer (Avanti Polar Lipids, 860462) standards, and normalization for 1-deoxydihydroceramides was done with the 13C-dihydroceramide-d7 standard. Homogenate aliquot of 50 µl was taken to determine tissue protein content using BCA protein assay (Lambda Biotech, G1002). The remaining homogenate was transferred to a 2 ml Eppendorf tube and 1 ml of −20 °C chloroform was added. Samples were vortex-mixed for 5 min and spun down for 5 min at 4 °C at 15,000*g*. The organic phase was collected and 2 µl of formic acid was added to the remaining polar phase which was re-extracted with 1 ml of −20 °C chloroform. Combined organic phases were dried and the pellet was resuspended in 100 µl of buffer containing 100% methanol, 1 mM ammonium formate and 0.2% formic acid. Data represents ion counts normalized by class-specific internal standards and tissue protein content, with stacked plots to represent acyl-chain distribution.

Gas chromatography–mass spectrometry

Quantification of polar metabolites was determined after derivatization with 2% (w/v) methoxyamine hydrochloride (Thermo Scientific) in pyridine (37 °C for 60 min) and with *N*-tertbutyldimethylsilyl-*N*-methyltrifluoroacetamide (MTBSTFA) with 1% *tert*-butyldimethylchlorosilane (tBDMS) (Regis Technologies) (37 °C for 30 min). Polar derivatives were analysed by GC–MS using a DB-35MS column (30 m × 0.25 mm internal diameter × 0.25 µm, Agilent J&W Scientific) installed in an Agilent 7890 A gas chromatograph interfaced with an Agilent 5975 C mass spectrometer as previously described[65]. Plasma glucose enrichment was determined using propionic anhydride derivatization as previously described[66]. Natural isotope abundance was corrected using in-house script[51].

Targeted sphingolipid quantification

Quantification of sphingolipid metabolites was determined using triple quadrupole liquid chromatography–mass spectrometry platform (Agilent 6460). Sphingolipid species were separated on a C8 column (Spectra 3 µm C8SR 150 × 3 mm inner diameter, Peeke Scientific) as previously described[67]. Mobile phase A was composed of 100% HPLC-grade water containing 2 mM ammonium formate and 0.2% formic acid and mobile phase B consisted of 100% methanol containing 0.2% formic acid and 1 mM ammonium formate. The flow rate was 0.5 ml min^{-1}. The gradient elution programme consisted of the following profile: 0 min, 82% B; 3 min, 82% B; 4 min, 90% B, 18 min, 99% B; 25 min, 99%, 27 min, 82% B, 30 min, 82% B. Column re-equilibration followed each sample and lasted 10 min. The capillary voltage was set to 3.5 kV, the drying gas temperature was 350 °C, the drying gas flow rate was 10 l min^{-1}, and the nebulizer pressure was 60 psi. Sphingolipid species were analysed by selective reaction monitoring (SRM) of the transition from precursor to product ions at associated optimized collision energies and fragmentor voltages (Supplementary Table 2). Quantification of sphingolipid species was performed using spiked-in deuterated standards.

High-resolution LC–MS/MS of polar metabolites

Around 10–20 mg of frozen tissue was extracted with 800 µl of −20 °C 5:3:2 acetonitrile: methanol:water solution spike with a known concentration of norvaline as an internal standard using the Precellys Evolution Homogenizer (Bertin Technologies)[68]. After extraction, a 50-µl aliquot was taken for protein quantification using BCA protein assay (Lambda Biotech, G1002), and the remaining extract was spun for 10 min at 21,000*g* at 4 °C. The supernatant was then transferred into a glass vial, and chromatographic separation and compound identification performed using Q Exactive Orbitrap MS with a Vanquish Flex Binary UHPLC system (ThermoFisher Scientific) on an iHILIC-(P) Classic, 150 mm by 2.1 mm, 5-mm particle, 200-Å (Hilicon) column at 45 °C. Chromatography was performed using a gradient of 20 mM ammonium carbonate, adjusted to pH 9.4 with 0.1% ammonium hydroxide (25%) solution (mobile phase A) and 100% acetonitrile (mobile phase B), both at a flow rate of 0.2 ml min^{-1}. The liquid chromatography gradient ran linearly from 80 to 20% B from 2 to 17 min and then from 20 to 80% B from 17 to 18 min and then held at 80% B from 18 to 25 min.

High-resolution LC–MS/MS of lipids

Liver samples were extracted in 400 µl of −20 °C methanol using the Precellys Evolution Homogenizer (Bertin Technologies) spiked with EquiSPLASH labelled standard (Avanti Polar Lipids, 330731) and norvaline. After extraction, 50 µl aliquot was taken to quantify protein content using BCA protein assay (Lambda Biotech, G1002), and to the remaining extract were added 400 µl of −20 °C chloroform and 400 µl of ice-cold water. After vortexing for 5 min, samples were spun for 5 min at 4 °C at 15,000*g*, and the organic phase was collected. Two microlitres of formic acid were added to the remaining polar phase which was re-extracted with 400 µl of −20 °C chloroform, samples were vortex-mixed, and spun as described above. Combined organic phases were dried and the pellet was resuspended in 100 µl of isopropanol.

Chromatographic separation and lipid species identification was performed using Q Exactive orbitrap mass spectrometer with a Vanquish Flex Binary

UHPLC system (Thermo Scientific) equipped with an Accucore C30, 150 × 2.1 mm, 2.6 µm particle (Thermo) column at 40 °C. Five microlitres of sample was injected. Chromatography was performed using a gradient of 40:60 v/v water: acetonitrile with 10 mM ammonium formate and 0.1% formic acid (mobile phase A) and 10:90 v/v acetonitrile: propan-2-ol with 10 mM ammonium formate and 0.1% formic acid (mobile phase B), both at a flow rate of 0.2 ml min^{-1}. The liquid chromatography gradient ran from 30% to 43% B from 3–8 min, then from 43% to 50% B from 8-9 min, then 50–90% B from 9–18 min, then 90–99% B from 18–26 min, then held at 99% B from 26–30 min, before returning to 30% B in 6 min and held for a further 4 min.

Lipids were analysed in positive mode using spray voltage 3.2 kV. Sweep gas flow was 1 arbitrary units, auxiliary gas flow 2 arbitrary units and sheath gas flow 40 arbitrary units, with a capillary temperature of 325 °C. Full mass spectrometry (scan range 200–2,000 *m*/*z*) was used at 70,000 resolution with 10^6 automatic gain control and a maximum injection time of 100 ms. Data dependent MS2 (Top 6) mode at 17,500 resolution with automatic gain control set at 10^5 with a maximum injection time of 50 ms was used. Data were analysed using EI-Maven software, and peaks normalized to Avanti EquiSPLASH internal standard. Lipid species specific fragments used for identification and quantification are presented in the Supplementary Table 3.

Plasma sphingoid base extraction, hydrolysis and LC–MS analysis

Plasma sphingolipids were processed as previously described with minor modifications[69]. In brief, 50 µl of plasma was mixed with 0.5 ml of methanol and spiked with internal standards, sphinganine-d7, sphingosine-d7 and deoxysphinganine-d3 (Avanti lipids). The samples were placed on a mixer for 1 h at 37 °C, centrifuged at 2,800*g* and the supernatant collected and acid hydrolysed overnight at 65 °C with 75 µl of methanolic HCl (1N HCl,10M H_2O in methanol). Next, 100 µl of 10 M KOH was added to neutralize. 625 µl of chloroform, 100 µl of 2N NH_4OH and 500 µl of alkaline water were added, samples vortex-mixed and centrifuged for 5 min at 16,000*g*. The lower organic phase was washed three times with alkaline water and dried under air. LC–MS analysis was performed on an Agilent 6460 QQQ LC–MS/MS. Metabolite separation was achieved with a C18 column (Luna 100

× 2.1 mm, 3 µm, Phenomenex). Mobile phase A was composed of a 60:40 ratio of methanol:water and mobile phase B consisted of 100% methanol with 0.1% formic acid with 5 mM ammonium formate added to both mobile phases. The gradient elution programme consisted of holding at 40% B for 0.5 min, linearly increasing to 100% B over 15 min, and maintaining it for 9 min, followed by re-equilibration to the initial condition for 10 min. The capillary voltage was set to 3.5 kV, the drying gas temperature was 350 °C, the drying gas flow rate was 10 l min^{-1}, and the nebulizer pressure was 60 psi. Sphingoid bases were analysed by SRM of the transition from precursor to product ions at associated optimized collision energies and fragmentor voltages[16]. Sphingoid bases were then quantified from spiked internal standards of known concentration.

Serine dehydratase activity assay

Frozen liver and kidney samples were extracted in an ice-cold buffer containing 50 mM KH_2PO_4, 1 mM Na_2EDTA, and 1mM DTT, pH 8.0 using glass homogenizer. Maximal enzyme activity was determined using coupled-enzyme reaction with lactate dehydrogenase (Sigma 10127230001) in the presence of 200 mM serine, 0.25 mM NADH, 0.17 mM pyridoxal phosphate, and 1 mM DTT for 3 min. Tissue homogenate protein quantification was subsequently determined using BCA protein assay (Lambda Biotech, G1002), and maximal enzyme activity expressed in international units (U) per mg of protein.

Gene expression analysis

RNA was extracted from ~20 mg of liver tissue using Direct-Zol RNA kit (Direct-Zol RNA Miniprep Plus kit, Zymo Research) according to the manufacturer's instructions. cDNA synthesis was performed using iScript Reverse Transcription Supermix for RT–PCR (iScript Reverse Transcription Supermix, Bio-Rad) according to the manufacturer's instructions using the following protocol: 5 min at 25 °C, 20 min 46 °C, 1 min 95 °C. PCR reactions were carried out using 96-well plates on an Applied Biosystems ViiA 7 Real-Time PCR System using the following parameters: 95 °C for 20 s, 40 cycles of 95 °C for 1 s, and 60 °C for 20 s. The final volume (10 µl) of PCR SYBR-Green reaction consisted of 5 µl fast SYBR-Green Master

Mix (Applied Biosystems), 2 µl cDNA, 1 µl of 5 µM forward and reverse primers, and 1 µl of water.

Primers used are as follows. 18s forward: AGTCCCTGCCCTTTGTACACA, 18s reverse: CGATCCGAGGGCCTCACTA; *Acc1* forward: AATGAACGTGCAATCCGATTTG, *Acc1* reverse: ACTCCACATTTGCGTAATTGTTG; *Acc2* forward: CGCTCACCAACAGTAAGGTGG, *Acc2* reverse: GCTTGGCAGGGAGTTCCTC; *Acly* forward: AATCCTGGCTAAAACCTCGCC, *Acly* reverse: GCATAGATGCACACGTAGAACT; actin forward: GGCTGTATTCCCCTCCATCG, actin reverse: CCAGTTGGTAACAATGCCATGT; *Aldh1l1* forward: AGCCACCTATGAGGGCATTC, *Aldh1l1* reverse: TGAGTGTCGAGTTGAAAAACGTC; *Aldh1l2* forward: ACCAGCCGGGTTTATTTCAAA, *Aldh1l2* reverse: ACTCCCACTACTCGGTGGC; *Dgat1* forward: CTGATCCTGAGTAATGCAAGGTT, *Dgat1* reverse: TGGATGCAATAATCACGCATGG; *Dgat2* forward: GCGCTACTTCCGAGACTACTT, *Dgat2* reverse: GGGCCTTATGCCAGGAAACT; *Dhcr7* forward: AGGCTGGATCTCAAGGACAAT, *Dhcr7* reverse: GCCAGACTAGCATGGCCTG; *Dhcr24* forward: CTCTGGGTGCGAGTGAAGG, *Dhcr24* reverse: TTCCCGGACCTGTTTCTGGAT; *Dld* forward: AGCTGGAGTCGTGTGTACC, *Dld* reverse: GAACCTATCACTGTCACGTCA; *Fasn* forward: GGAGGTGGTGATAGCCGGTAT, *Fasn* reverse: TGGGTAATCCATAGAGCCCAG; *Fdft1* forward: GTTTGAAGACCCCATAGTTGGTG, *Fdft1* reverse: CACATCTACGTTCTCTGGCTTAG; *Fdps* forward: GGAGGTCCTAGAGTACAATGCC, *Fdps* reverse: AAGCCTGGAGCAGTTCTACAC; *Ggps1* forward: TTCACAGGCATTTAATCACTGGC, *Ggps1* reverse: ACCACGTCGGAGCTTTGAAC; *Gldc* forward: CTCCTGCCCAGACACGAT, *Gldc* reverse: GGGACCGTCTTCTCGATGAG; *Gpat1* forward:

CTTGGCCGATGTAAACACACC, *Gpat1* reverse: CTTCCGGCTCATAAGGCTCTC; *Gpat4* forward: TCAAAGAAATTCGTCGAAGTGGT, *Gpat4* reverse: CCTTTCCGGCAAAAGTAGAAGAT; *Hmgcs1* forward: AACTGGTGCAGAAATCTCTAGC, *Hmgcs1* reverse: GGTTGAATAGCTCAGAACTAGCC; *Hmgcr* forward: AGCTTGCCCGAATTGTATGTG, *Hmgcr* reverse: TCTGTTGTGAACCATGTGACTTC; *Lss* forward: TCGTGGGGGACCCTATAAAAC, *Lss* reverse: CGTCCTCCGCTTGATAATAAGTC; *Mthfd1* forward: CTCCTGTCCCAAGTGACATTG, *Mthfd1* reverse: TAGCCTTCGTTTCCCCGTACA; *Mthfd2* forward: AGTGCGAAATGAAGCCGTTG, *Mthfd2* reverse: GACTGGCGGGATTGTCACC; *Mthfd1l* forward: GCATGGCCTTACCCTTCAGAT, *Mthfd1l* reverse: GTACGAGCTTCCCCAGATTGA; *Mthfd2l* forward: AAGGACGTTGATGGATTTCACAT, *Mthfd2l* reverse: GATGATTTCCCAAACGGCACT; *Mthfr* forward: AGATGAGGCGCAGAATGGAC, *Mthfr* reverse: CATCCGGTCAAACCTGGAGAT; *Mtr* forward: TCCTCCTCGGCCTATCTTTATTT, *Mtr* reverse: GGTCCGAATGAGACACGCT; *Mvk* forward: GGTGTGGTCGGAACTTCCC, *Mvk* reverse: CCTTGAGCGGGTTGGAGAC; *Mvd* forward: ATGGCCTCAGAAAAGCCTCAG, *Mvd* reverse: TGGTCGTTTTTAGCTGGTCCT; *Pmvk* forward: CCTATGGGGCTGTGATACAGA, *Pmvk* reverse: TCTCCGTGGTTCTCAATGACC; *Psat* forward: CAGTGGAGCGCCAGAATAGAA, *Psat* reverse: CCTGTGCCCCTTCAAGGA; *Psph* forward: TGAGTACGCAGGTTTTGATGAG, *Psph* reverse: TGAGTACGCAGGTTTTGATGAG; *Phgdh* forward: ATGGCCTTCGCAAATCTGC, *Phgdh* reverse: AGTTCAGCTATCAGCTCCTCC; *Scd1* forward: TTCTTGCGATACACTCTGGTGC, *Scd2* reverse: CGGGATTGAATGTTCTTGTCGT; *Scd2* forward: GATCTCTGGCGCTTACTCAGC, *Scd2* reverse:

CTCCCCAGTGGTGAGAACTC; *Sds* forward: GAAGACCCCACTTCGTGACAG, *Sds* reverse: TCTTGCAGAGATGCCCAATGC; *Shmt1* forward: CAGGGCTCTGCTTGATGCAC, *Shmt1* reverse: CGTAACGCGCTCTTGTCAC; *Shmt2* forward: TGGCAAGAGATACTACGGAGG, *Shmt2* reverse: AGATCCGCTTGACATCAGACA; *Sqle* forward: ATAAGAAATGCGGGGATGTCAC, *Sqle* reverse: ATATCCGAGAAGGCAGCGAAC; *Srebp1a* forward: TAGTCCGAAGCCGGGTGGGCGCCGG, *Srebp1a* reverse: GATGTCGTTCAAAACCGCTGTGTGTC; *Srebp1c* forward: AAGCAAATCACTGAAGGACCTGG, *Srebp1c* reverse: AAAGACAAGCTACTCTGGGAG; *Srebp2* forward: GGATCCTCCCAAAGAAGGAG, *Srebp2* reverse: TTCCTCAGAACGCCAGACTT; *Tyms* forward: GGAAGGGTGTTTTGGAGGAGT, *Tyms* reverse: GCTGTCCAGAAAATCTCGGGA.

Western blotting

To compare tissue protein levels, ~20 mg of tissue was homogenized in RIPA buffer supplemented with 5 mM EDTA solution (Thermo Scientific), protease inhibitor cocktail (cOmplete, Roche) and phosphatase inhibitor cocktail (PhosSTOP, Roche), and placed on ice for 30 min. Tissue lysate was centrifuged at 13,000*g* for 10 min at 4 °C, and the supernatant was stored at −80 °C. Homogenate protein content was determined using the bicinchoninic acid assay (Thermo Scientific). Protein samples were prepared in a Laemmli buffer (NuPAGE LDS Sample Buffer, Life Technologies), and were run a 4–15% precast gel (Mini-PROTEAN TGX, Bio-Rad) for 2 h at constant 100 V and transferred on a polyvinylidenedifluoride (PVDF) membrane for 2 h at constant 250 mA in an ice-chilled transfer tank. The membrane was blocked with 5% milk and incubated overnight at 4 °C with a primary antibody against ACLY (Cell Signaling 13390, 1:1,000), ACC (Cell Signaling 3662, 1:2,000), p-AKT Ser473 (Cell Signaling 9271, 1:1,000), p-AKT Ser308 (Cell Signaling 9275, 1:1,000), AKT (Cell Signaling 75692, 1:1,000), SCD1 (Cell Signaling 2794, 1:1,000), GAPDH (Cell Signaling 5174, 1:4,000), and vinculin (Cell Signaling 4650, 1:1,000). After washing

with TBS-T, membranes were incubated with a horseradish peroxidase-conjugated secondary antibody (Cell Signaling 7074, 1:5,000) for 1 h at room temperature and incubated with enhanced chemiluminescence liquid (Clarity Western ECL Substrate, Bio-Rad) for 5 min. Densitometry quantification was performed using Image Lab software (Bio-Rad). For raw scans with molecular weight markers, see Supplementary Fig. 1.

Statistical analysis

Data are expressed as mean ± s.e.m. unless stated otherwise. Statistical analysis was performed with Prism software (GraphPad Prism 9.3.1) using two-sided independent *t*-test to compare two groups, one-way ANOVA with Fisher's least significant difference post hoc test to compare more than two groups, two-way ANOVA with Fisher's least significant difference post hoc test to compare two-factor study design, and PERMANOVA analysis to explore RPCA plots. For all tests, $P < 0.05$ was considered significant. All data points in the manuscript represent individual biological replicates. No statistical methods were used to predetermine sample size.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Source data for microbiome algorithms and immunoblots are provided as Supplementary Information. The whole-microbiome genome sequencing raw data was uploaded to Qiita[53], where we followed the default processing workflow. High-resolution and targeted mass spectrometry data data is available at the NIH Common Fund's National Metabolomics Data Repository (NMDR) website, the Metabolomics Workbench, https://www.metabolomicsworkbench.org where it has been assigned Project ID M8JD81 (https://doi.org/10.21228/M8JD81). The data can be accessed directly via the Project ID M8JD81. Additional data that support the findings are available from the corresponding author upon reasonable request.

References

1. Feldman, E. L. et al. Diabetic neuropathy. *Nat. Rev. Dis. Primers* **5**, 41 (2019).

2. Forbes, J. M. & Cooper, M. E. Mechanisms of diabetic complications. *Physiol. Rev.* **93**, 137–188 (2013).

3. Newgard, C. B. Interplay between lipids and branched-chain amino acids in development of insulin resistance. *Cell Metab.* **15**, 606–614 (2012).

4. Yang, Q., Vijayakumar, A. & Kahn, B. B. Metabolites as regulators of insulin sensitivity and metabolism. *Nat. Rev. Mol. Cell Biol.* **19**, 654–672 (2018).

5. Spiegelman, B. M. & Flier, J. S. Obesity and the regulation of energy balance. *Cell* **104**, 531–543 (2001).

6. Zochodne, D. W. Sensory neurodegeneration in diabetes: beyond glucotoxicity. *Int. Rev. Neurobiol.* **127**, 151–180 (2016).

7. Eid, S. et al. New insights into the mechanisms of diabetic complications: role of lipids and lipid metabolism. *Diabetologia* **62**, 1539–1549 (2019).

8. Felig, P., Marliss, E. & Cahill, G. F. Jr. Plasma amino acid levels and insulin secretion in obesity. *New Engl. J. Med.* **281**, 811–816 (1969).

9. Huffman, K. M. et al. Relationships between circulating metabolic intermediates and insulin action in overweight to obese, inactive men and women. *Diabetes Care* **32**, 1678–1683 (2009).

10. Newgard, C. B. et al. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab.* **9**, 311–326 (2009).

11. Bervoets, L. et al. Metabolic profiling of type 1 diabetes mellitus in children and adolescents: a case-control study. *Diabetol. Metab. Syndr.* **9**, 48 (2017).

12. Thalacker-Mercer, A. E. et al. BMI, RQ, diabetes, and sex affect the relationships between amino acids and clamp measures of insulin action in humans. *Diabetes* **63**, 791–800 (2014).

13. Gaggini, M. et al. Altered amino acid concentrations in NAFLD: impact of obesity and insulin resistance. *Hepatology* **67**, 145–158 (2018).

14. Laidlaw, S. A. et al. Patterns of fasting plasma amino acid levels in chronic renal insufficiency: results from the feasibility phase of the Modification of Diet in Renal Disease Study. *Am. J. Kidney Dis.* **23**, 504–513 (1994).

15. Fridman, V. et al. Altered plasma serine and 1-deoxydihydroceramide profiles are associated with diabetic neuropathy in type 2 diabetes and obesity. *J. Diabetes Complications* **35**, 107852 (2021).

16. Gantner, M. L. et al. Serine and lipid metabolism in macular disease and peripheral neuropathy. *New Engl. J. Med.* **381**, 1422–1433 (2019).

17. Mardinoglu, A. et al. Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nat. Commun.* **5**, 3083 (2014).

18. Lowry, M., Hall, D. E. & Brosnan, J. T. Serine synthesis in rat kidney: studies with perfused kidney and cortical tubules. *Am. J. Physiol.* **250**, F649–F658 (1986).

19. LeBoeuf, S. E. et al. Activation of oxidative stress response in cancer generates a druggable dependency on exogenous non-essential amino acids. *Cell Metab.* **31**, 339–350.e334 (2020).

20. Maddocks, O. D. et al. Serine starvation induces stress and p53-dependent metabolic remodelling in cancer cells. *Nature* **493**, 542–546 (2013).

21. Muthusamy, T. et al. Serine restriction alters sphingolipid diversity to constrain tumour growth. *Nature* **586**, 790–795 (2020).

22. Gheller, B. J. et al. Extracellular serine and glycine are required for mouse and human skeletal muscle stem and progenitor cell function. *Mol. Metab.* **43**, 101106 (2021).

23. Ridaura, V. K. et al. Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* **341**, 1241214 (2013).

24. Wallace, M. et al. Enzyme promiscuity drives branched-chain fatty acid synthesis in adipose tissues. *Nat. Chem. Biol.* **14**, 1021–1031 (2018).

25. Eade, K. et al. Serine biosynthesis defect due to haploinsufficiency of PHGDH causes retinal disease. *Nature Metab.* **3**, 366–377 (2021).

26. Bonelli, R. et al. Systemic lipid dysregulation is a risk factor for macular neurodegenerative disease. *Sci. Rep.* **10**, 12165 (2020).

27. Le Douce, J. et al. Impairment of glycolysis-derived l-serine production in astrocytes contributes to cognitive deficits in Alzheimer's disease. *Cell Metab.* **31**, 503–517.e508 (2020).

28. Scerri, T. S. et al. Genome-wide analyses identify common variants associated with macular telangiectasia type 2. *Nat. Genet.* **49**, 559–567 (2017).

29. Kennedy, J. M. & Zochodne, D. W. Experimental diabetic neuropathy with spontaneous recovery: is there irreparable damage? *Diabetes* **54**, 830–837 (2005).

30. Esaki, K. et al. l-Serine deficiency elicits intracellular accumulation of cytotoxic deoxysphingolipids and lipid body formation. *J. Biol. Chem.* **290**, 14595–14609 (2015).

31. Lone, M. A., Santos, T., Alecu, I., Silva, L. C. & Hornemann, T. 1-Deoxysphingolipids. *Biochim. Biophys. Acta* **1864**, 512–521 (2019).

32. Chaurasia, B. et al. Adipocyte ceramides regulate subcutaneous adipose browning, inflammation, and metabolism. *Cell Metab.* **24**, 820–834 (2016).

33. Hines, T. J. et al. Precision mouse models of *Yars*/dominant intermediate Charcot–Marie–Tooth disease type C and *Sptlc1*/hereditary sensory and autonomic neuropathy type 1. *J. Anat.* **241**, 1169–1185 (2022).

34. Hill, R. Z., Morita, T., Brem, R. B. & Bautista, D. M. S1PR3 mediates itch and pain via distinct TRP channel-dependent pathways. *J. Neurosci.* **38**, 7833–7843 (2018).

35. Johnson, V. J., He, Q., Osuchowski, M. F. & Sharma, R. P. Disruption of sphingolipid homeostasis by myriocin, a mycotoxin, reduces thymic and splenic T-lymphocyte populations. *Toxicology* **201**, 67–75 (2004).

36. Lowry, M., Hall, D. E., Hall, M. S. & Brosnan, J. T. Renal metabolism of amino acids in vivo: studies on serine and glycine fluxes. *Am. J. Physiol.* **252**, F304–F309 (1987).

37. Leung, K. Y. et al. Regulation of glycine metabolism by the glycine cleavage system and conjugation pathway in mouse models of non-ketotic hyperglycinemia. *J. Inherit. Metab. Dis.* **43**, 1186–1198 (2020).

38. Floegel, A. et al. Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. *Diabetes* **62**, 639–648 (2013).

39. Fridman, V. et al. Randomized trial of l-serine in patients with hereditary sensory and autonomic neuropathy type 1. *Neurology* **92**, e359–e370 (2019).

40. Othman, A. et al. Lowering plasma 1-deoxysphingolipids improves neuropathy in diabetic rats. *Diabetes* **64**, 1035–1045 (2015).

41. Jolivalt, C. G. et al. B vitamins alleviate indices of neuropathic pain in diabetic rats. *Eur. J. Pharmacol.* **612**, 41–47 (2009).

42. Anderson, N. J., King, M. R., Delbruck, L. & Jolivalt, C. G. Role of insulin signaling impairment, adiponectin and dyslipidemia in peripheral and central neuropathy in mice. *Dis. Models Mech.* **7**, 625–633 (2014).

43. Parks, B. W. et al. Genetic architecture of insulin resistance in the mouse. *Cell Metab.* **21**, 334–347 (2015).

44. Nicholson, A. et al. Diet-induced obesity in two C57BL/6 substrains with intact or mutant nicotinamide nucleotide transhydrogenase (*Nnt*) gene. *Obesity* **18**, 1902–1905 (2010).

45. Clark, A. J. et al. An iPSC model of hereditary sensory neuropathy-1 reveals l-serine-responsive deficits in neuronal ganglioside composition and axoglial interactions. *Cell Rep. Med.* **2**, 100345 (2021).

46. Truman, J. P. et al. Sphingosine kinase 1 downregulation is required for adaptation to serine deprivation. *FASEB J.* **35**, e21284 (2021).

47. Mohassel, P. et al. Childhood amyotrophic lateral sclerosis caused by excess sphingolipid synthesis. *Nat. Med.* **27**, 1197–1204 (2021).

48. Clarke, B. A. et al. The *Ormdl* genes regulate the sphingolipid synthesis pathway to ensure proper myelination and neurologic function in mice. *eLife* **8**, e51067 (2019).

49. Bonelli, R. et al. Identification of genetic factors influencing metabolic dysregulation and retinal support for MacTel, a retinal disorder. *Commun. Biol.* **4**, 274 (2021).

50. Zhang, Y., Huo, M., Zhou, J. & Xie, S. PKSolver: an add-in program for pharmacokinetic and pharmacodynamic data analysis in Microsoft Excel. *Comput. Methods Programs Biomed.* **99**, 306–314 (2010).

51. Fernandez, C. A., Des Rosiers, C., Previs, S. F., David, F. & Brunengraber, H. Correction of ^{13}C mass isotopomer distributions for natural stable isotope abundance. *J. Mass Spectrom.* **31**, 255–262 (1996).

52. Heikkinen, S., Argmann, C. A., Champy, M. F. & Auwerx, J. Evaluation of glucose homeostasis. *Curr. Protoc. Mol. Biol.* https://doi.org/10.1002/0471142727.mb29b03s77 (2007).

53. Gonzalez, A. et al. Qiita: rapid, web-enabled microbiome meta-analysis. *Nat. Methods* **15**, 796–798 (2018).

54. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

55. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

56. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

57. Zhu, Q. et al. Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea. *Nat. Commun.* **10**, 5477 (2019).

58. Bolyen, E. et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).

59. Faith, D. P. Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* **61**, 1–10 (1992).

60. Martino, C. et al. A novel sparse compositional technique reveals microbial perturbations. *mSystems* **4**, e00016–e00019 (2019).

61. Anderson, M. J. *Permutational Multivariate Analysis of Variance (PERMANOVA)* https://doi.org/10.1002/9781118445112.stat07841 (2017).

62. Stan Development Team. *Stan Modeling Language Users Guide and Reference Manual*, Version 2.27 https://mc-stan.org/docs/2_27/stan-users-guide/ (2021).

63. Caspi, R. et al. The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Res.* **46**, D633–d639 (2018).

64. Jolivalt, C. G. et al. Peripheral neuropathy in mouse models of diabetes. *Curr. Protoc. Mouse Biol.* **6**, 223–255 (2016).

65. Cordes, T. & Metallo, C. M. Quantifying intermediary metabolism and lipogenesis in cultured mammalian cells using stable isotope tracing and mass spectrometry. *Methods Mol. Biol.* **1978**, 219–241 (2019).

66. Antoniewicz, M. R., Kelleher, J. K. & Stephanopoulos, G. Measuring deuterium enrichment of glucose hydrogen atoms by gas chromatography/mass spectrometry. *Anal. Chem.* **83**, 3211–3216 (2011).

67. Bielawski, J. et al. Comprehensive quantitative analysis of bioactive sphingolipids by high-performance liquid chromatography-tandem mass spectrometry. *Methods Mol. Biol.* **579**, 443–467 (2009).

68. Lim, E. W. et al. Progressive alterations in amino acid and lipid metabolism correlate with peripheral neuropathy in Polg(D257A) mice. *Sci. Adv.* **7**, eabj4077 (2021).

69. Othman, A. et al. Plasma 1-deoxysphingolipids are predictive biomarkers for type 2 diabetes mellitus. *BMJ Op. Diabetes Res. Care* **3**, e000073 (2015).

Acknowledgements

We thank all members of the Metallo laboratory for helpful discussions as well as M. Friedlander, E. Holmes and A. Tayerani for their input. This work was supported by the NIH (R01CA234245 to C.M.M., NIDDK MICROMouse Funding DK076169 to M.K.H., and R01AG065993 to A.C.), a Camille and Henry Dreyfus Teacher–Scholar Award (to C.M.M.), Lowy Medical Research Institute (to C.M.M.) and the American Heart Association (18CDA34110292 to A.C.). This work was also supported by the NIDDK-funded San Diego Digestive Diseases Research Center grant (P30 DK120515).

Author information

Authors and Affiliations

1. Molecular and Cell Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA, USA

   Michal K. Handzlik, Jivani M. Gengatharan, Grace H. McGregor, Courtney R. Green, Patrick Tseng & Christian M. Metallo

2. Department of Bioengineering, University of California San Diego, La Jolla, CA, USA

   Michal K. Handzlik, Jivani M. Gengatharan, Grace H. McGregor, Ana M. Moreno, Courtney R. Green, Prashant Mali, Rob Knight & Christian M. Metallo

3. Department of Pathology, School of Medicine, University of California San Diego, La Jolla, CA, USA

   Katie E. Frizzi, Lucie S. Guernsey & Nigel A. Calcutt

4. Department of Pediatrics, School of Medicine, University of California San Diego, La Jolla, CA, USA

   Cameron Martino, Gibraan Rahman, Antonio Gonzalez & Rob Knight

5. Bioinformatics and Systems Biology Program, University of California San Diego, La Jolla, CA, USA

   Cameron Martino, Gibraan Rahman & Rob Knight

6. Center for Microbiome Innovation, University of California San Diego, La Jolla, CA, USA

   Cameron Martino & Rob Knight

7. Regulatory Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA, USA

   Terry Lin & Satchidananda Panda

8. Scripps Research, La Jolla, CA, USA

   Yoichiro Ideguchi

9. Lowy Medical Research Institute, La Jolla, CA, USA

   Regis J. Fallon & Marin L. Gantner

10. Department of Nutrition and Integrative Physiology, University of Utah, Salt Lake City, UT, USA

    Amandine Chaix

11. School of Agriculture and Food Science, University College Dublin, Dublin, Ireland

    Martina Wallace

Contributions

M.K.H. and C.M.M. designed the study. M.K.H., J.M.G., A.M.M., Y.I., K.E.F., L.S.G., P.T. and R.J.F. performed behavioural assays, GTTs and ITTs. M.K.H., G.H.M. and C.R.G. generated and analysed stable isotope tracer and targeted metabolomics and lipidomics data. C.M., G.R. and A.G. performed microbiome analysis, including phylogenetic alpha and beta-diversity and pathway analysis. T.L. and A.C. performed faecal bomb calorimetry experiments. S.P., M.W., P.M., R.K., M.L.G., N.A.C. and C.M.M. guided experimental design and analysis. M.K.H. and C.M.M. wrote the manuscript with input from all authors.

Corresponding author

Correspondence to Christian M. Metallo.

Ethics declarations

Competing interests

C.M.M. is a scientific advisor for Faeth Therapeutics. The other authors declare no competing interests.

Peer review

Peer review information

Nature thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Serine and glycine metabolism in diabetes.

(**a**) Levels of glycine, serine, and methionine in WT and BKS-*db/db* mice in the kidney (n = 6 per group). (**b**) Levels of glycine, serine, and methionine in WT and BKS-*db/db* mice in the inguinal adipose tissue (iWAT) (n = 6 per group). (**c**) Plasma metabolite levels in WT and BKS-*db/db* mice (n = 6 per group). (**d**) mRNA expression of kidney enzymes regulating serine, glycine, and one-carbon metabolism in WT (n = 6) and BKS-*db/db* mice (n = 5). (**e**) Serine tolerance test (STT) in C57BL/6J mice after an overnight fast (n = 4 per dose). (**f**) Tissue serine labeling fraction in WT mice 15 min after [U-$^{13}C_3$]serine administration via oral gavage (n = 4 per tissues) after an overnight fast. (**g**) Tissue citrate labeling fraction in WT mice 15 min after [U-$^{13}C_3$]serine administration via oral gavage (n = 4 per tissues) after an overnight fast. Data are mean ± standard error of mean (SEM) and were

analyzed using a two-sided independent t-test (**a–d**). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$ vs. WT group.

Extended Data Fig. 2 Serine tolerance test in type 1 and 2 diabetic mice.

(**a**) Serine dehydratase activity determined in the liver and kidney in C57BL/6J mice in the fed state (n = 5 per tissues). (**b**) Serum insulin in WT and BKS-*db/db* mice quantified after a 6-h fast (n = 6 per group). (**c**) Serum glucagon in WT (n = 5) and BKS-*db/db* (n = 6) mice quantified after a 6-h fast. (**d**) Plasma glucose pharmacokinetics and area under curve (AUC_{GLC}) in WT and BKS-*db/db* mice following GTT/STT with glucose dosed at 2 g/kg (n = 6 per group) after an overnight fast. (**e**) Plasma glycine pharmacokinetics and area under curve (AUC_{GLY}) in WT and BKS-*db/db* mice following GTT/STT with glucose dosed at 2 g/kg (n = 6 per group) after an overnight fast. (**f**) Serine tolerance test (STT) and area under curve (AUC_{SER}) in WT and BKS-*db/db* mice (n = 6 per group) dosed with serine (400 mg/kg) only after an overnight fast. (**g**) Fasting (6-h) plasma insulin level in C57BL/6J mice treated with vehicle or streptozotocin (STZ) 1 week after injection (n = 9 per group). (**h**) Fasting (6-h) blood glucose in C57BL/6J mice treated with vehicle or streptozotocin (STZ) 1 week after injection (n = 9 per group). (**i**) Body fat mass content in C57BL/6J mice treated with vehicle or streptozotocin (STZ) 1 week after injection (n = 9 per group). (**j**) Fasting (6-h) plasma amino acid concentration in C57BL/6J mice treated with vehicle (n = 9) or streptozotocin (STZ, n = 10) 1 week after injection. (**k**) Plasma glucose pharmacokinetics and area under curve (AUC_{GLC}) after an overnight fast during GTT/STT in C57BL/6J mice treated with vehicle (n = 7) or streptozotocin (STZ, n = 6) 2 weeks after injections with glucose dosed at 2 g/kg. (**l**) Thermal latency in 14-week-old WT and BKS-*db/db* mice (n = 6 per group). (**m**) Tactile sensing in 14-week-old WT and BKS-*db/db* mice (n = 6 per group). (**n**) Motor nerve conduction velocity in 14-week-old WT and BKS-*db/db* mice (n = 6 per group). Data are mean ± standard error of mean (SEM) and were analyzed using a two-sided independent t-test (b–n) and a two-way ANOVA with Fisher's LSD *post hoc* test (e–f, k time-course experiments).

Extended Data Fig. 3 Dietary serine restriction and metabolic homeostasis.

(*a*) Plasma amino acid level in the fed state in mice fed with a low fat diet (LFD, n = 5), serine/glycine-free LFD (-SG LFD, n = 8), high-fat diet (HFD, n = 8), and serine/glycine free HFD (-SG HFD, n = 8) for 18 weeks. (*b*) Liver metabolite content in mice fed with a low fat diet (LFD, n = 10), serine/glycine-free LFD (-SG LFD, n = 10), high-fat diet (HFD, n = 10), and serine/glycine free HFD (-SG HFD, n = 10) for 18 weeks. (*c*) Body weight gain 18 weeks after dietary intervention (n = 13 per group). (*d*) Food intake in response to an 18-week dietary feeding in mice with a low fat diet (LFD, n = 10), serine/glycine-free LFD (-SG LFD, n = 12), high fat diet (HFD, n = 13), and serine/glycine-free HFD (-SG HFD, n = = 13). (*e*) Calorie intake in response to an 18-week dietary feeding in mice with a low fat diet (LFD, n = 10), serine/glycine-free LFD (-SG LFD, n = 12), high fat diet (HFD, n = 13), and serine/glycine-free HFD (-SG HFD, n = 13). (*f*) Calorie absorption in response to an 18-week dietary feeding in mice with a low fat diet (LFD, n = 10), serine/glycine-free LFD (-SG LFD, n = 12), high fat diet (HFD, n = 13), and serine/glycine-free HFD (-SG HFD, n = 13). (*g*) Water intake in response to an 18-week dietary feeding in mice with a low fat diet (LFD, n = 10), serine/glycine-free LFD (-SG LFD, n = 12), high fat diet (HFD, n = 13), and serine/glycine-free HFD (-SG HFD, n = 13). (*h*) Physical activity in response to an 18-week dietary feeding in mice with a low fat diet (LFD, n = 10), serine/glycine-free LFD (-SG LFD, n = 12), high fat diet (HFD, n = 13), and serine/glycine-free HFD (-SG HFD, n = 13). (*i*) Representative images and quantification of adipocyte size in response to an 18-week dietary feeding in mice with a low fat diet (LFD, n = 10), serine/glycine-free LFD (-SG LFD, n = = 12), high fat diet (HFD, n = 13), and serine/glycine-free HFD (-SG HFD, n = 13). (*j*) I.P. glucose tolerance test 18 weeks after feeding with a low fat diet (LFD) or high fat diet (HFD) replete or deficient in serine and glycine (n = 13 per group). (*k*) I.P. insulin tolerance test 18 weeks after feeding with a low fat diet (LFD) or high fat diet (HFD) replete or deficient in serine and glycine (n = 15 per group). (*l*) Respiratory exchange ratio time course 18 weeks after feeding mice with a low fat diet (LFD, n = 10), serine/glycine-free LFD (-SG LFD, n = 12), high fat diet (HFD, n = 13), and serine/glycine-free HFD (-SG HFD, n = = 13). Data are mean ± standard error of mean (SEM) and were analyzed using a two-way

ANOVA with Fisher's LSD *post hoc* test (**a–k**). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$ vs. LFD. # $P < 0.05$ vs. HFD group.

Extended Data Fig. 4 The impact of dietary serine restriction on microbiome composition.

(*a*) Robust principal-component analysis of microbiome beta-diversity 18 weeks after feeding mice with a low fat diet (LFD, n = 10), serine/glycine-free LFD (-SG LFD, n = 10), high fat diet (HFD, n = 9), and serine/glycine-free HFD (-SG HFD, n = 9). (*b*) Log-ratio of species with complete vs. incomplete pathways for L-serine biosynthesis 18 weeks after feeding mice with a low fat diet (LFD, n = 10), serine/glycine-free LFD (-SG LFD, n = 10), high fat diet (HFD, n = 9), and serine/glycine-free HFD (-SG HFD, n = 9). (*c*) Log-ratio of species with complete vs. incomplete pathways for glycine cleavage 18 weeks after feeding mice with a low fat diet (LFD, n = 10), serine/glycine-free LFD (-SG LFD, n = 10), high fat diet (HFD, n = 9), and serine/glycine-free HFD (-SG HFD, n = 9). (*d*) Log fold-change of microbiome species 18 weeks after feeding mice with a low fat diet (LFD, n = 10), serine/glycine-free LFD (-SG LFD, n = 10), high fat diet (HFD, n = 9), and serine/glycine-free HFD (-SG HFD, n = 9). Data are presented as a minimum/maximum (**b–c**) and were analyzed using a PERMANOVA test (*a*) and a two-way ANOVA with Fisher's LSD *post hoc* test (**b–c**).

Extended Data Fig. 5 Dietary serine restriction remodels fatty acid and cholesterol synthesis.

(*a*) Plasma heavy water (D_2O) enrichment in mice fed a with low fat diet (LFD, n = 3), serine/glycine-free LFD (-SG LFD, n = 4), high fat diet (HFD, n = 5), and serine/glycine-free HFD (-SG HFD, n = 5). (*b*) *De novo* cholesterol synthesis in mice fed with LFD (n = 3), -SG LFD (n = 4), HFD (n = 5), and -SG HFD (n = 5) for 18 weeks. (*c*) Hepatic protein expression of ATP-citrate lyase (ACLY), acetyl-CoA carboxylase (ACC), and steroyl-CoA desaturase 1 (SCD1), P-Akt^{S473}, and P-Akt^{T308} (n = 3 per group). (*d*) Hepatic mRNA expression of ATP-citrate lyase (*Acly*), acetyl-CoA carboxylase (*Acc2*), and steroyl-CoA desaturase 1 (*Scd1*) in mice fed with a low fat diet (LFD, n = 8), serine/glycine-free LFD (-SG LFD, n = 11), high fat diet (HFD, n = 13), and serine/glycine-free HFD (-SG HFD, n = 10). (*e*)

Hepatic mRNA expression of enzymes involved in fatty acid and cholesterol synthesis in mice fed with a low fat diet (LFD, n = 8), serine/glycine-free LFD (-SG LFD, n = 11), high fat diet (HFD, n = 13), and serine/glycine-free HFD (-SG HFD, n = 10). Data are mean ± standard error of mean (SEM) and were analyzed using a two-way ANOVA with Fisher's LSD *post hoc* test (**a**–**e**). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$ vs. LFD. # $P < 0.05$ vs. HFD group.

Extended Data Fig. 6 Myriocin remodels obesity and sphingolipid metabolism in serine-associated peripheral neuropathy.

(*a*) Thermal latency quantification in C57BL/6J mice fed with either a control diet (n = 10) or serine/glycine-free how (-Ser/Gly, n = 10) for 12 months. (*b*) Representative images of intraepidermal nerve fiber (IENF) density and quantification of IENF in paw skin in mice fed with either a control (n = 2) or serine/glycine free diet (n = 3) for 12 months. PGP9.5 staining of IENF is shown by arrows. (*c*) Body weight time course in mice fed with a low fat diet + vehicle (LFD+Veh, n = 10), LFD + 0.3 mg/kg myriocin (LFD+Myr, n = 10), serine/glycine-free LFD + vehicle (-SG LFD+Veh, n = 10), -SG LFD + myriocin (-SG LFD+Myr, n = 9), high fat diet + vehicle (HFD+Veh, n = 10), HFD + myriocin (HFD+Myr, n = 10), serine/glycine-free HFD + vehicle (-SG HFD+Veh, n = 10), and -SG HFD + myriocin (-SG HFD+Myr, n = 9). (*d*) Body weight gain 6 months after dietary intervention in mice fed with a low fat diet + vehicle (LFD+Veh, n = 10), LFD + 0.3 mg/kg myriocin (LFD+Myr, n = 10), serine/glycine-free LFD + vehicle (-SG LFD+Veh, n = 10), -SG LFD + myriocin (-SG LFD+Myr, n = 9), high fat diet + vehicle (HFD+Veh, n = 10), HFD + myriocin (HFD+Myr, n = 10), serine/glycine-free HFD + vehicle (-SG HFD+Veh, n = 10), and -SG HFD + myriocin (-SG HFD+Myr, n = 9). (*e*) Adipose tissue size 6 months after dietary intervention in mice fed with a low fat diet + vehicle (LFD+Veh, n = 9), LFD + 0.3 mg/kg myriocin (LFD+Myr, n = 10), serine/glycine-free LFD + vehicle (-SG LFD+Veh, n = 10), -SG LFD + myriocin (-SG LFD+Myr, n = 9), high fat diet + vehicle (HFD+Veh, n = 10), HFD + myriocin (HFD+Myr, n = 10), serine/glycine-free HFD + vehicle (-SG HFD+Veh, n = 10), and -SG HFD + myriocin (-SG HFD+Myr, n = 9). (*f*) Plasma amino acid concentration 6 months after dietary intervention in mice

fed a with low fat diet + vehicle (LFD+Veh, n = 10), LFD + 0.3 mg/kg myriocin (LFD+Myr, n = 10), serine/glycine-free LFD + vehicle (-SG LFD+Veh, n = 10), -SG LFD + myriocin (-SG LFD+Myr, n = 9), high fat diet + vehicle (HFD+Veh, n = 10), HFD + myriocin (HFD+Myr, n = 10), serine/glycine-free HFD + vehicle (-SG HFD+Veh, n = 10), and -SG HFD + myriocin (-SG HFD+Myr, n = 9). (*g*) Stack plot of liver ceramides in mice fed with a low fat diet + vehicle (LFD+Veh, n = 10), LFD + 0.3 mg/kg myriocin (LFD+Myr, n = 10), serine/glycine-free LFD + vehicle (-SG LFD+Veh, n = 10), -SG LFD + myriocin (-SG LFD+Myr, n = 9), high fat diet + vehicle (HFD+Veh, n = 10), HFD + myriocin (HFD+Myr, n = 10), serine/glycine-free HFD + vehicle (-SG HFD+Veh, n = 10), and -SG HFD + myriocin (-SG HFD+Myr, n = 9). (*h*) Stack plot of sciatic nerve ceramides in mice fed with a low fat diet + vehicle (LFD+Veh, n = 10), LFD + 0.3 mg/kg myriocin (LFD+Myr, n = 10), serine/glycine-free LFD + vehicle (-SG LFD+Veh, n = 10), -SG LFD + myriocin (-SG LFD+Myr, n = 9), high fat diet + vehicle (HFD+Veh, n = 10), HFD + myriocin (HFD+Myr, n = 10), serine/glycine-free HFD + vehicle (-SG HFD+Veh, n = 10), and -SG HFD + myriocin (-SG HFD+Myr, n = 9). (*i*) Stack plot of sciatic nerve deoxydihydroceramide (deoxyDHCer) in mice fed a with low fat diet + vehicle (LFD+Veh, n = 10), LFD + 0.3 mg/kg myriocin (LFD+Myr, n = 10), serine/glycine-free LFD + vehicle (-SG LFD+Veh, n = 10), -SG LFD + myriocin (-SG LFD+Myr, n = 9), high fat diet + vehicle (HFD+Veh, n = 10), HFD + myriocin (HFD+Myr, n = 10), serine/glycine-free HFD + vehicle (-SG HFD+Veh, n = 10), and -SG HFD + myriocin (-SG HFD+Myr, n = 9). Data are mean ± standard error of mean (SEM) and were analyzed using a two-way ANOVA with Fisher's LSD *post hoc* test (**a** and **c**) and a one-way ANOVA with Fisher's LSD *post hoc* test (**d–i**). Statistical analyses in g-i were performed using summed sphingolipid abundances.

Extended Data Fig. 7 Myriocin improves innervation in the context of serine deficiency and alters the hepatic lipidome.

(*a*) Representative images of intraepidermal nerve fiber (IENF) density in mice fed a with low fat diet + vehicle (LFD+Veh, n = 10), high fat diet + vehicle (HFD+Veh, n = 7), serine/glycine free HFD + vehicle (-SG HFD+Veh, n = 12), and -SG HFD + myriocin (-SG HFD+Myr, n = 8). (*b*) Representative images and quantification of corneal nerves in mice fed a

with low fat diet + vehicle (LFD+Veh, n = 12), high fat diet + vehicle (HFD+Veh, n = 12), serine/glycine-free HFD + vehicle (-SG HFD+Veh, n = 12), and -SG HFD + myriocin (-SG HFD+Myr, n = 11). (*c*) Tactile sensing in mice fed with a low fat diet + vehicle (LFD+Veh, n = 12), high fat diet + vehicle (HFD+Veh, n = 12), serine/glycine-free HFD + vehicle (-SG HFD+Veh, n = 12), and -SG HFD + myriocin (-SG HFD+Myr, n = 11). (*d*) Motor nerve conduction velocity in mice fed with a low fat diet + vehicle (LFD+Veh, n = 12), high fat diet + vehicle (HFD+Veh, n = 12), serine/glycine-free HFD + vehicle (-SG HFD+Veh, n = 12), and -SG HFD + myriocin (-SG HFD+Myr, n = 11). (*e*) Liver high-resolution lipidomics analysis in the fed state in mice fed with a low fat diet + vehicle (LFD+Veh, n = 12), high fat diet + vehicle (HFD+Veh, n = 12), serine/glycine-free HFD + vehicle (-SG HFD+Veh, n = 12), and -SG HFD + myriocin (-SG HFD+Myr, n = 11). Data are mean ± standard error of mean (SEM) and were analyzed using a one-way ANOVA with Fisher's LSD *post hoc* test (**b–e**).

Extended Data Fig. 8 Myriocin slows the progression of peripheral neuropathy in BKS-*db/db* mice.

a) Hydrolyzed plasma deoxysphinganine (deoxySA) concentration in 16-week-old WT and BKS-*db/db* mice (n = 6 per group). **b**) Hepatic targeted sphingolipid determination 16-week-old WT and BKS-*db/db* mice (n = 6 per group). **c**) Thermal latency time-course in BKS-*db/db* mice treated with either vehicle (n = 10) or myriocin (0.3 mg/kg every other day, n = 9) for 8 weeks. **d**) Tactile sensing in BKS-*db/db* mice treated with either vehicle (n=10) or myriocin (n = 9) for 8 weeks. **e**) Intraepidermal nerve fiber density in BKS-*db/db* mice treated with either vehicle (n = 10) or myriocin (n = 9) for 8 weeks. **f**) Body weight time-course in BKS-*db/db* mice treated with either vehicle (n = 10) or myriocin (n = 9) for 8 weeks. **g**) Fasting (6-h) blood glucose time-course in BKS-*db/db* mice treated with either vehicle (n = 10) or myriocin (n = 9) for 8 weeks. **h**) Fasting (6-h) plasma serine time-course in BKS-*db/db* mice treated with either vehicle (n = 10) or myriocin (n = 8) for 8 weeks. **i**) Summed hepatic sphingolipids in BKS-*db/db* mice treated with either vehicle (n = 10) or myriocin (n = 9) for 8 weeks. **j**) Summed paw skin sphingolipids in BKS-*db/db* mice treated with either vehicle (n = 10) or myriocin (n = 9) for 8 weeks. **k**) Stack plot of liver deoxydihydroceramide (deoxyDHCer) species in BKS-*db/db* mice treated

with either vehicle (n = 10) or myriocin (n = 9) for 8 weeks. **l**) Stack plot of paw skin deoxydihydroceramide (deoxyDHCer) species in BKS-*db/db* mice treated with either vehicle (n = 10) or myriocin (n = 9) for 8 weeks. m) Stack plot of liver sphingomyelin (SM) species in BKS-*db/db* mice treated with either vehicle (n = 10) or myriocin (n = 9) for 8 weeks. **n**) Stack plot of paw skin sphingomyelin (SM) species in BKS-*db/db* mice treated with either vehicle (n = 10) or myriocin (n = 9) for 8 weeks. Data are mean ± standard error of mean (SEM) and were analyzed using a two-sided independent t-test (**a–b**,**d–e**, **i–n**) and a two-way ANOVA with Fisher's LSD *post hoc* test (**c**,**f–h**). Statistical analyses in k-n were performed using summed sphingolipid abundances.

Extended Data Fig. 9 Dietary serine supplementation mitigates 1-deoxysphingolipid in BKS-*db/db* liver and skin.

a) Body weight time-course in BKS-*db/db* mice fed with either a control (n = 8) or serine supplemented diet (n = 9) for 8 weeks. **b**) Fasting (6-h) blood glucose time-course in BKS-*db/db* fed with either a control (n = 8) or serine supplemented diet (n = 9) for 8 weeks. **c**) Summed hepatic sphingolipids in BKS-*db/db* mice fed with either a control or serine supplemented diet for 8 weeks (n = 8 per group). **d**) Summed paw skin sphingolipids in BKS-*db/db* mice fed with either a control or serine supplemented diet for 8 weeks (n = 8 per group). **e**) Levels of sphingomyelin species (SM) in the liver in BKS-*db/db* mice fed with either a control or serine supplemented diet for 8 weeks (n = 8 per group). **f**) Levels of sphingomyelin species (SM) in the paw skin in BKS-*db/db* mice fed with either a control or serine supplemented diet for 8 weeks (n = 8 per group). Data are mean ± standard error of mean (SEM) and were analyzed with a two-way ANOVA with Fisher's LSD *post hoc* test (**a–b**) and a two-sided independent t-test (**c–f**). Statistical analyses in e-f were performed using summed abundances.

Supplementary information

Supplementary Figure 1

Supplementary Fig. 1 contains the raw western blot images.

Reporting Summary

Supplementary Table 1

Dietary composition.

Supplementary Table 2

LC–MS/MS transitions.

Supplementary Table 3

High-resolution mass spectrometry transitions.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Handzlik, M.K., Gengatharan, J.M., Frizzi, K.E. *et al.* Insulin-regulated serine and lipid metabolism drive peripheral neuropathy. *Nature* **614**, 118–124 (2023). https://doi.org/10.1038/s41586-022-05637-6

- Received: 20 December 2021
- Accepted: 07 December 2022
- Published: 25 January 2023
- Issue Date: 02 February 2023
- DOI: https://doi.org/10.1038/s41586-022-05637-6

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

Amino acid slows nerve damage from diabetes, in mouse study

- Nick Petrić Howe
- Shamini Bundell

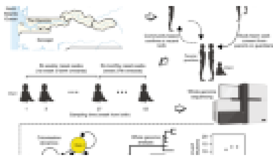
Nature Nature Podcast 25 Jan 2023

Serine deficiency causes complications in diabetes

- Thorsten Hornemann

Nature News & Views 25 Jan 2023

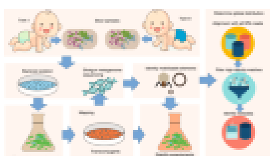
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05637-6



Within-host microevolution of Streptococcus pneumoniae is rapid and adaptive during natural colonisation

10 July 2020

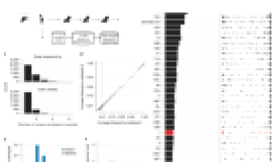
Chrispin Chaguza, Madikay Senghore, … Brenda A. Kwambana-Adams



Experimental support for multidrug resistance transfer potential in the preterm infant gut microbiota

01 July 2019

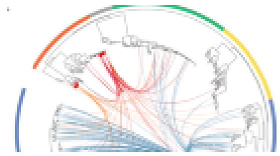
Mari Hagbø, Anuradha Ravi, … Knut Rudi



Pneumococcal within-host diversity during colonization, transmission and treatment

10 October 2022

Gerry Tonkin-Hill, Clare Ling, … Stephen D. Bentley



[Strain-level characterization of broad host range mobile genetic elements transferring antibiotic resistance from the human microbiome]

17 March 2022

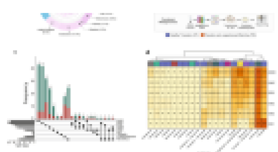
Samuel C. Forster, Junyan Liu, … Trevor D. Lawley



[Population genomics of Escherichia coli in livestock-keeping households across a rapidly developing urban landscape]

14 March 2022

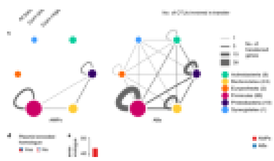
Dishon M. Muloi, Bryan A. Wee, … Eric M. Fèvre



[Impact of international travel and diarrhea on gut microbiome and resistome dynamics]

05 December 2022

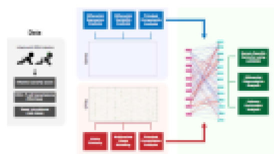
Manish Boolchandani, Kevin S. Blake, … Gautam Dantas



[Phylogenetic barriers to horizontal transfer of antimicrobial peptide resistance genes in the human gut microbiota]

17 December 2018

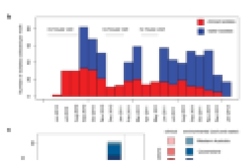
Bálint Kintses, Orsolya Méhi, … Csaba Pál



[Microbiome-Transcriptome Interactions Related to Severity of Respiratory Syncytial Virus Infection]

25 September 2019

Abhijeet R. Sonawane, Liang Tian, … Yang-Yu Liu



[Genetic variation associated with infection and the environment in the accidental pathogen Burkholderia pseudomallei]

22 November 2019

Claire Chewapreecha, Alison E. Mather, … Sharon J. Peacock

- Article
- Open Access
- Published: 18 January 2023

The person-to-person transmission landscape of the gut and oral microbiomes

- Mireia Valles-Colomer ,
- Aitor Blanco-Míguez,
- Paolo Manghi,
- Francesco Asnicar,
- Leonard Dubois,
- Davide Golzato,
- Federica Armanini,
- Fabio Cumbo,
- Kun D. Huang,
- Serena Manara,
- Giulia Masetti,
- Federica Pinto,
- Elisa Piperni,
- Michal Punčochář,
- Liviana Ricci,
- Moreno Zolfo,
- Olivia Farrant,
- Adriana Goncalves,
- Marta Selma-Royo,
- Ana G. Binetti,
- Jimmy E. Becerra,
- Bei Han,
- John Lusingu,
- John Amuasi,
- Loredana Amoroso,
- Alessia Visconti,
- Claire M. Steves,
- Mario Falchi,
- Michele Filosi,
- Adrian Tett,

- Anna Last,
- Qian Xu,
- Nan Qin,
- Huanlong Qin,
- Jürgen May,
- Daniel Eibach,
- Maria Valeria Corrias,
- Mirco Ponzoni,
- Edoardo Pasolli,
- Tim D. Spector,
- Enrico Domenici,
- Maria Carmen Collado &
- …
- Nicola Segata

Show authors

Nature **volume 614**, pages 125–135 (2023)

- 52k Accesses
- 1 Citations
- 943 Altmetric
- Metrics details

Abstract

The human microbiome is an integral component of the human body and a co-determinant of several health conditions[1,2]. However, the extent to which interpersonal relations shape the individual genetic makeup of the microbiome and its transmission within and across populations remains largely unknown[3,4]. Here, capitalizing on more than 9,700 human metagenomes and computational strain-level profiling, we detected extensive bacterial strain sharing across individuals (more than 10 million instances) with distinct mother-to-infant, intra-household and intra-

population transmission patterns. Mother-to-infant gut microbiome transmission was considerable and stable during infancy (around 50% of the same strains among shared species (strain-sharing rate)) and remained detectable at older ages. By contrast, the transmission of the oral microbiome occurred largely horizontally and was enhanced by the duration of cohabitation. There was substantial strain sharing among cohabiting individuals, with 12% and 32% median strain-sharing rates for the gut and oral microbiomes, and time since cohabitation affected strain sharing more than age or genetics did. Bacterial strain sharing additionally recapitulated host population structures better than species-level profiles did. Finally, distinct taxa appeared as efficient spreaders across transmission modes and were associated with different predicted bacterial phenotypes linked with out-of-host survival capabilities. The extent of microorganism transmission that we describe underscores its relevance in human microbiome studies[5], especially those on non-infectious, microbiome-associated diseases.

Main

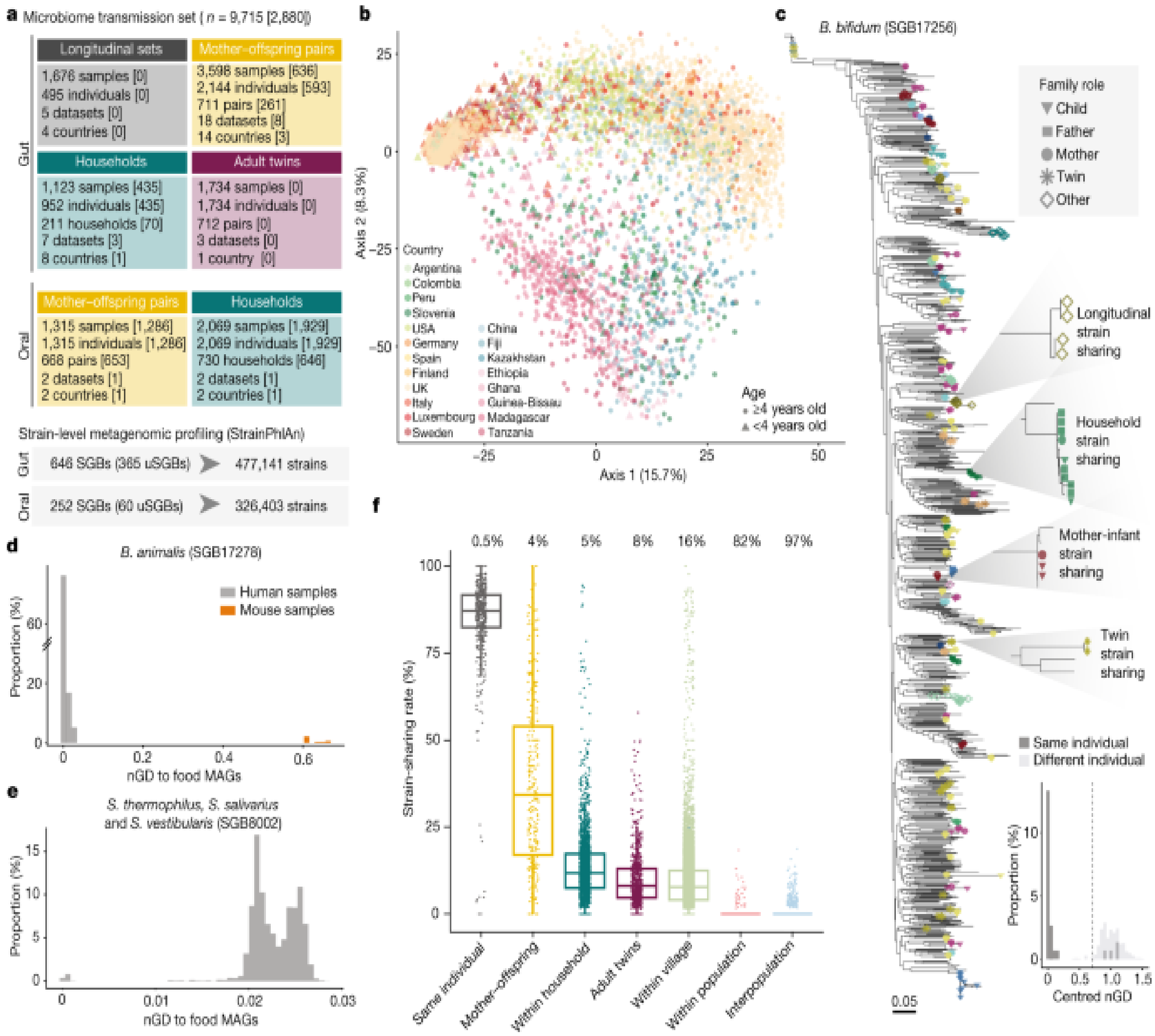
Our genome is inherited from our parents and remains stable over our lifetime, with limited accumulation of nucleotide variations. By contrast, the genetic makeup of our microorganism complement (the human microbiome) is seeded at birth and changes over time, displaying both high temporal variability and personalization[6,7]. Factors including diet and lifestyle are well known to modulate the composition of the human microbiome[1,2,8], but as very few members of the microbiome can thrive outside the human body, most microorganisms must be acquired from other individuals[3,4]. Indeed, colonization of the human gut by microorganisms is largely seeded by maternal transmission[9,10,11,12,13,14], but maternal seeding alone cannot account for the large diversity of microorganisms found in adults. How members of the microbiome are acquired and transmitted by individuals and spread in populations, and how this shapes the personal microbiome genetic makeup remain largely unexplored—especially in humans[15,16]—with only preliminary findings to date[11,17]. So far, research has been hindered by the limited number and size of accurately designed studies, and by the difficulties in consistently and comprehensively profiling microorganism conspecific strains—that is, genetic variants within species.

Strains are the individual-specific building blocks of the human microbiome[18,19]. They can be highly genomically and functionally divergent within a species, and their profiling is a necessary prerequisite to distinguish transmission of microorganisms from microbiome convergence towards an overlapping set of species. Identifying the features of microbiome transmission will advance our understanding of the complexity of the human microbiome, and can help address the ‘communicable’ factor that microbiome transmission adds to diseases and conditions currently considered non-communicable[5]. Here, we characterize and quantify the patterns of person-to-person microbiome strain sharing across multiple scenarios to provide a comprehensive description of the microbiome transmission landscape.

Profiling microbiome transmission

To unravel the modes of person-to-person microbiome transmission we performed an integrative analysis on a large set of metagenomic datasets[2,9,10,12,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34] with known family relationships ($n = 31$) that were analysed using improved strain-level profiling metagenomic tools (Methods). Eight of these datasets were newly sequenced in the context of this study from different geographical areas and host lifestyles in America (Argentina, Colombia and the USA), Africa (Guinea-Bissau), Asia (China) and Europe (Italy). Three other studies[9,34] in Africa (Ghana and Tanzania) and Europe (Italy) were further expanded here for a total of 978 stool and 1,929 saliva samples (Supplementary Tables 1 and 2). This collection comprises 9,715 microbiome samples (7,646 stool and 2,069 saliva) and curated host information, enabling the assessment of transmission across mother–infant pairs, household members, adult twin pairs, villages and populations. Although the 31 datasets differ in size, with human metagenomes from 20 different countries in five continents and representing diverse host lifestyles (Fig. 1a,b, Extended Data Fig. 1a and Supplementary Table 2), the integrated set facilitates the identification of person-to-person microbiome transmission patterns at the global level.

Fig. 1: A metagenomic framework to survey person-to-person microbiome strain transmission.



a, Overview of the study and dataset based on the SGB framework (Methods). Numbers in square brackets are the number of units sequenced in this study. **b**, Overall species-level structure of the gut samples (principal component analysis on Aitchison distance, one random sample per individual, $n = 4{,}840$). Samples are coloured by country and shapes indicate age. **c**, Phylogeny of *B. bifidum* (SGB17256) (Methods), a low-prevalence highly transmitted species (Supplementary Table 9), showing the genetic diversity of strains and the shared strains between samples of the same individual and between different individuals. One example of strain sharing is highlighted for each relationship type. Tree leaves involved in strain-sharing instances are coloured by dataset (Extended Data Fig. 1b) and their shapes reflect kinship. Bottom, the distribution of pairwise centred nGDs of the species in individuals sampled at two time points (less than six months

apart, 'same individual') and in unrelated individuals ('different individual'; Extended Data Fig. 3 and Methods), confirming the suitability of the methodology to infer strain identity. **d**,**e**, The distribution of pairwise nGDs between *B. animalis* (SGB17278) (**d**) and *S. thermophilus*, *S. salivarius* and *S. vestibularis* (SGB8002) (**e**) strains reconstructed from human gut metagenomes or mouse samples and MAGs reconstructed from fermented food[40]. The presence of *B. animalis* in humans is associated with the consumption of commercial dietary products (Extended Data Fig. 4a), whereas only a subset of *S. thermophilus*, *S. salivarius* and *S. vestibularis* strains is associated with fermented food intake (Extended Data Fig. 4b). **f**, Person-to-person strain-sharing rates (number of shared strains/number of shared SGBs × 100%) across relationship types. All comparisons are statistically significant (Kruskal–Wallis test, $n = 26{,}218$, $\chi^2 = 11{,}420$, $P < 2.2 \times 10^{-16}$, post hoc Dunn tests, $P_{adj} < 0.05$; Supplementary Table 7). In box plots, box edges delineate lower and upper quartiles, the centre line represents the median and whiskers extend to 1.5 times the interquartile range (IQR). The number along the top is the percentage of pairs between which no strain-sharing event was detected.

Microorganism strain transmission inference via metagenomics exploits the validated assumption that strains usually persist within an individual's gut over periods of at least a few months but are rarely found in unrelated individuals unless direct or indirect transmission has occurred[19,35,36,37,38]. Here, we first improved our strain-level profiling methodology[39] (Methods), and then further refined strain tracking with operational species-specific definitions of strain identity (Extended Data Fig. 2). Strain boundaries were set by identifying the normalized phylogenetic distance (nGD) thresholds that best separated same-individual longitudinal strain retention from unrelated individual nGD distributions in more than 1,500 longitudinal samples from 4 countries[20,22,27,28,31] (Youden's index allowing <5% potential false positives—that is, same strain shared by unrelated individuals; permutation ANOVA (PERMANOVA), $n \geq 50$ pairs, $R^2 = 0.75$ to 1%, $P < 0.001$; Fig. 1c, Extended Data Fig. 3, Supplementary Table 3 and Methods). Such nGD-based thresholds perform well with phylogenies built with the rather low average coverage that is typical for most detectable species in metagenomic samples (mean coverage = 7.2×) and with limited lengths of the concatenated marker gene alignments (mean trimmed

alignment length = 74,348 nucleotides (nt)). In addition, our approach exploits the information on evolutionary models that is provided by phylogenetic trees that is not available when considering raw single-nucleotide variation (SNV) rates or genetic similarity.

Microbiome profiling was also expanded to 1,022 not yet cultured and unnamed species (referred to as unknown species-level genome bins (uSGBs)), complementing the 1,730 species with cultured representatives (known species-level genome bins (kSGBs)) defined in a repository of more than 214,000 metagenome-assembled genomes (MAGs) and around 138,000 available isolate genomes[39]. uSGBs constitute 37% of all detected species-level genome bins (SGBs) and were found to be highly prevalent (86% of gut and 100% of oral metagenomes, with 17% and 10% median relative abundance, respectively), especially in gut metagenomes from non-westernized communities (99% prevalence, with 42% median relative abundance overall; Methods). Strain sharing was assessed by profiling in each sample the dominant strain of SGBs found with at least 10% prevalence and in at least 20 samples of at least one cohort, for a total of 646 SGBs in gut metagenomes (Supplementary Table 4) and 252 SGBs in oral metagenomes (Supplementary Table 5), with 24 SGBs profiled in both environments. The developed computational methodology is publicly available for strain-transmission inference from any metagenomic dataset (Methods and Code availability).

As a case in point, *Bifidobacterium bifidum* (SGB17256)—one of the 646 gut SGBs assessed for transmission—was successfully profiled in 1,298 gut microbiome samples (17% of total stool samples). We detected the same *B. bifidum* strain in 87% of pairs of samples from the same individual collected up to six months apart, with nGD between strains following a clear bimodal distribution (the first peak at phylogenetic distance close to zero indicating shared strains) (Fig. 1c). Overall, 13,278 instances of inter-individual shared *B. bifidum* strains were identified between the vast majority of mothers and their offspring (proportion of strain-sharing events detected over potential transmissions—that is, SGB transmissibility = 0.93; Methods) as well as among household members (SGB transmissibility = 0.73).

Even though disentangling direct transmission from indirect acquisition or co-acquisition is possible only with longitudinal sampling or in specific

settings (for example, mother to newborn), we minimized the chances of detecting strain sharing resulting from co-acquisition from common dietary sources by identifying and discarding in each SGB those strains with high similarity (≤0.0015 SNV rate) to MAGs or isolate genomes of microorganisms obtained from commercial fermented foods[40] (Methods). Because food microbiomes remain poorly investigated, other strains or species might originate from food sources even though food-to-gut colonization is regarded as rare[40]. This filtering resulted in the exclusion from the downstream analysis of most *Bifidobacterium animalis* (SGB17278) strains (278 strains, 94% of the total; Fig. 1d, Extended Data Fig. 4a, Supplementary Table 6 and Methods) in gut samples, supporting its putative origin from commercial dietary products[20]. Indeed, more than 98% of excluded samples were from westernized datasets, whereas only 6 strains were detected in non-westernized datasets (less than 0.07% of non-westernized samples), from locations where commercial probiotics are less available. Following the same criterion, 540 strains being phylogenetically close to MAGs of food origin were excluded from 7 other SGBs, including *Streptococcus thermophilus, S. salivarius* and *S. vestibularis* (SGB8002) (19 strains excluded; Fig. 1e, Extended Data Fig. 4b and Supplementary Table 6). Overall, after these exclusions, we detected around 6.35 million instances of strain sharing between different individuals in gut samples and around 4.91 million in oral samples.

Overview of gut microbiome transmission

We first assessed general gut microbiome strain-sharing patterns across human relationships, defining person-to-person strain-sharing rates as the number of strains shared between two individuals normalized by the number of SGBs profiled in common (out of the 646 SGBs profiled at strain level; Methods). Strains were confirmed to be highly persistent in subjects sampled less than six months apart[20,22,27,28,31] (median 87% strain-sharing rate), with as little as 0.5% of individuals displaying no longitudinal overlap in the detected strains—potentially owing to the occurrence of unreported perturbations or sample mislabelling. The highest person-to-person strain-sharing rates were detected between cohabiting mothers and their 0- to 3-year-old offspring (median of 34% strain-sharing rate), followed by individuals 4 years of age and older in the same household (12%), non-

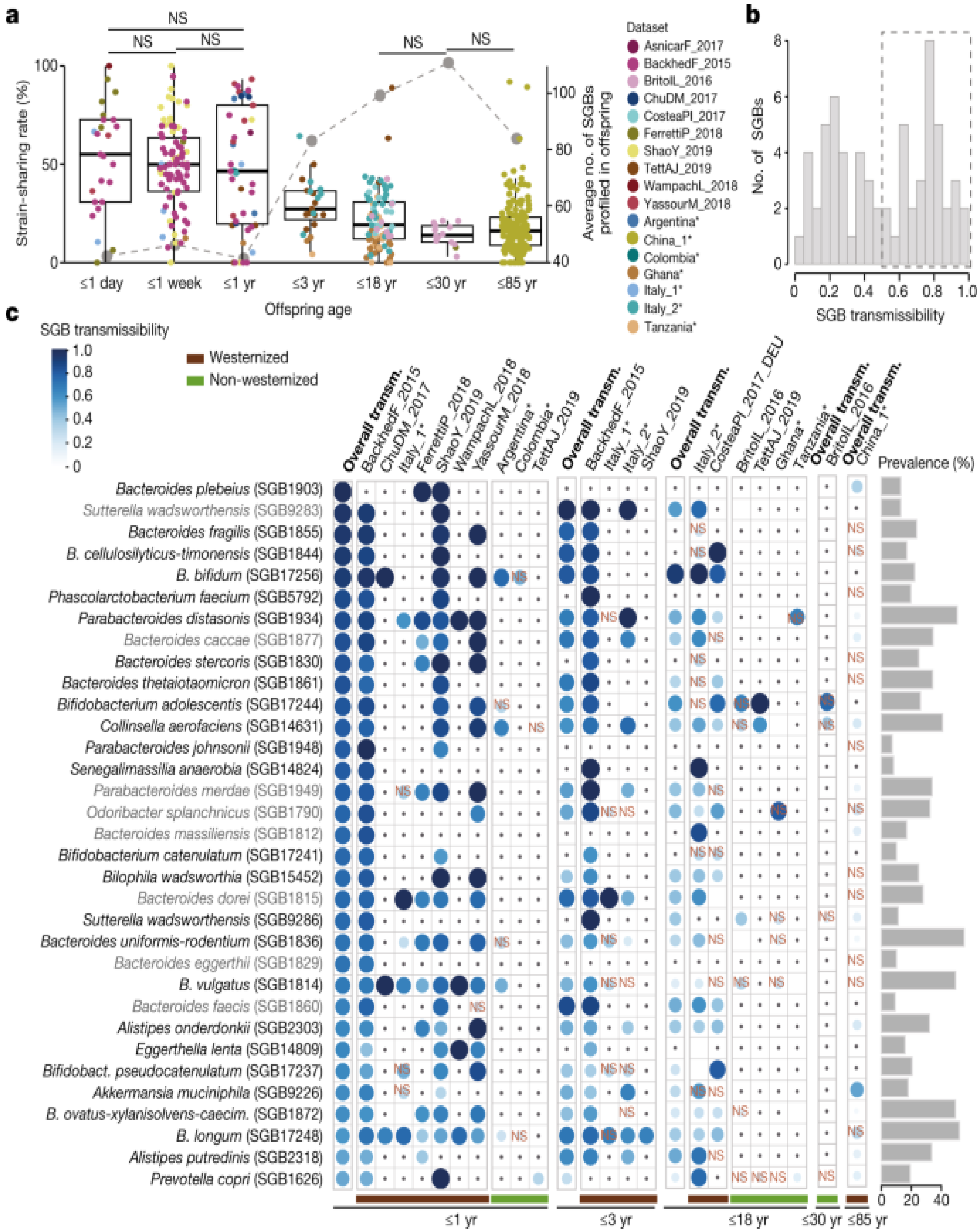
cohabiting adult twins (8%), and non-cohabiting adults in the same village (8%). Whereas strain sharing between adult twins might in part result from persisting shared maternal transmission, strain sharing among individuals in the same village is probably the result of horizontal transmission through physical interaction and the shared environment. By contrast, non-cohabiting individuals in different villages of the same and of different population-specific study cohorts (hereafter 'populations') displayed minimal strain-sharing rates (median 0%) (Kruskal–Wallis test, $n = 26{,}218$, $\chi^2 = 11{,}420$, $P < 2.2 \times 10^{-16}$, post hoc Dunn tests, adjusted P value (P_{adj}) < 0.05; Fig. 1f and Supplementary Table 7). This highly significant pattern is confirmed by the percentage of individuals not sharing a single detectable strain: whereas only 4% of mother–offspring pairs had no detected strain-sharing event, no strains were shared by 82% of pairs with no obvious person-to-person contact in the same population, and by up to 97% of individuals in different populations (Fig. 1f). Person-to-person strain sharing thus follows a social distance-based gradient across shared environments and kinship that is notably stronger than that observed by species-level microorganism divergence (beta diversity indices, Kruskal–Wallis tests with post hoc Dunn tests, $P_{adj} < 0.05$; Extended Data Fig. 4b and Supplementary Table 8). Overall, our integrated analysis highlights the relevance of direct person-to-person interaction and social-interaction networks in shaping the gut microbiome of single individuals.

Extensive mother–offspring transmission

Mother-to-offspring microbiome transmission has been described[9,10,11,29,32,41], and our expanded sample set (3,598 samples from 711 mother–offspring pairs, including 636 novel stool samples; Fig. 1a) enabled further generalization of the previously reported patterns. We found a remarkable negative correlation between the strain-sharing rate and the age of the offspring (Spearman's test, $n = 448$, $\rho = -0.52$, $P < 2.2 \times 10^{-16}$; Kruskal–Wallis test, $\chi^2 = 156$, $P < 2.2 \times 10^{-16}$; Fig. 2a) despite the increase on the number of mother–offspring shared species with offspring age (median = 17 shared species in the first year of life, 37 up to 3 years of age, and 57 up to 18 years of age), suggesting the accumulation of species putatively originating from other sources by the offspring. During the first

year of life, infants shared with their mothers half of the strains of the species found in both the infant and the mother microbiomes (strain-sharing rate) and 16% of the strains detected in the infants putatively originated from the mother (Extended Data Fig. 6a and Supplementary Table 10), with only slight non-significant reductions in strain-sharing rates after the first few days[9,12] (65%, 50% and 47% median strain-sharing rates at 1 day, 1 week, and 1 year, respectively; post hoc Dunn tests, $P_{adj} \geq 0.05$, Supplementary Table 10). In concordance with the reduced post-weaning physical intimacy and the infant's expanding motor activities[42], strain sharing then decreased to 27% at 1–3 years of age (Fig. 2a). Mother–offspring strain-sharing rates stabilized after 3 years of age (19% for up to 18 years of age and 14% for up to 30 years of age; Fig. 2a), approaching those observed between household members (12%; Fig. 1f). Whereas ample strain sharing at birth confirms the substantial extent of maternal microbiome seeding of the infant's gut, strain sharing remained significant in senior individuals (50–85 years of age), with non-cohabiting mother–offspring pairs still sharing significantly more strains than with unrelated mothers (16% versus 8%; Wilcoxon rank-sum test, $n = 17{,}177$, $r = 0.09$, $P = 4.1 \times 10^{-35}$; Extended Data Fig. 6b). This may be the result of the combined effect of long-lasting maternal microorganism imprinting at birth and strain transmission driven by shared social environments later in life.

Fig. 2: Mother–offspring sharing of the gut microbiome.



a, Mother–offspring strain-sharing rates (left axis; box plots and non-grey dots) decrease, whereas species richness (right axis; grey dots) in offspring increases, as a function of offspring age (17 datasets in 14 countries). The median number of SGBs profiled by StrainPhlAn in the offspring is used as

a proxy for richness (right axis). Kruskal–Wallis test, $n = 448$, $\chi^2 = 156$, $P < 2.2 \times 10^{-16}$, post hoc Dunn tests; NS, not significant ($P_{adj} \geq 0.05$); all other comparisons are significant (Supplementary Table 10). In box plots, box edges delineate lower and upper quartiles, the centre line represents the median and whiskers extend to 1.5 times the IQR. Novel datasets from the present study are highlighted with asterisks. **b**, The distribution of mother–infant SGB transmissibility in the first year of life. **c**, A panel of 33 SGBs that are highly maternally transmitted during their first year of life (SGB transmissibility >0.5 and significantly higher mother–infant transmissibility than unrelated mother–infant transmissibility; Methods) showing their transmissibility (transm.) in specific datasets (separated by westernized-lifestyle status) and in other age categories. NS, non-significant SGB transmissibility in the category (χ^2 test on the number of transmitted and non-transmitted SGBs between mother–offspring pairs and between unrelated mother and offspring pairs, Supplementary Table 16). Only comparisons with at least three possible transmissions (species shared by at least three mother–offspring pairs) are shown; comparisons with less than three possible transmissions are marked with a dot. Prevalence is defined as the percentage of mother–offspring samples in which the SGB was detected. Novel datasets from the present study are highlighted with asterisks. SGB names in grey use a strain identity threshold of 5% inter-individual nGD (Supplementary Table 4). *B. cellulosilyticus-timonensis*, *Bacteroides cellulosilyticus* and *Bacteroides timonensis*; *Bacteroides uniformis-rodentium*, *Bacteroides uniformis* and *Bacteroides rodentium*; *B. pseudocatenulatum*, *Bifidobacterium pseudocatenulatum*; *B. ovatus-xylanisolvens-caecim.*, *Bacteroides caecimuris*.

Potential effectors of maternal gut microbiome transmission include lifestyle and mode of delivery[14,29]. Although the newly sequenced non-westernized populations reinforced the well-documented westernization-associated reduction in microorganism diversity[43,44,45] both in mothers (Wilcoxon rank-sum test, $n = 721$, $r = -0.37$, $P = 7.4 \times 10^{-24}$) and their offspring ($P_{adj} < 0.05$, Extended Data Fig. 6c and Supplementary Table 11), we noticed no differential mother–offspring strain-sharing rates in most age categories (Wilcoxon rank-sum tests, $P_{adj} \geq 0.05$ for all age categories except for 3–18 years of age; Supplementary Table 12). Indeed, similar numbers of strains were maternally transmitted in westernized and non-westernized

communities (Wilcoxon rank-sum tests, $P_{adj} \geq 0.05$ for all age categories except for 3–18 years; Supplementary Table 13). The high microbiome diversity in non-westernized populations thus does not seem to be maintained by maternal transmission of microbiome strains but might be gained by closer interaction with more individuals. By contrast, we did confirm an association between mode of delivery and mother–offspring strain sharing early in life: vaginally delivered infants (up to 1 year of age) displayed significantly higher strain-sharing rates with their mothers (Wilcoxon rank-sum tests, $P_{adj} < 0.05$; Extended Data Fig. 6d and Supplementary Table 14). However, paralleling the age-associated decreased influence of mode of delivery on the infants' microbiome[46], no difference was detected after 3 years of age ($n = 56$, $r = 0.2$, $P_{adj} = 0.18$; Supplementary Table 14). Therefore, whereas vaginal delivery provides evident gut microbiome imprinting via maternal transmission early in life, lifestyle differences—including divergent hygiene and built-environment sanitation levels—do not substantially affect microbiome transmission rates.

Transmission from mothers to offspring (defined on offspring of up to 1 year of age—before the reduction in strain sharing; Fig. 2a) varied largely among species (Fig. 2b), but SGB transmissibility was rather consistent across datasets (pairwise Spearman's tests, $\rho = 0.59$–0.83, $P_{adj} < 0.05$; Supplementary Table 15), revealing species transmissibility as a specific trait of microorganisms. All highly transmitted SGBs (51% SGBs, transmissibility greater than 0.5 and significantly higher mother–infant transmissibility than unrelated mother–infant transmissibility; Methods) across 10 datasets belonged to characterized species (kSGBs) (Chi-squared tests, $n = 33$, $P_{adj} < 0.05$; Fig. 2c and Supplementary Table 16), mostly of the genera *Bacteroides* and *Bifidobacterium* ($n = 16$ (48%) and $n = 5$ (15%) SGBs, respectively; Fig. 2c). As a case in point, *Bacteroides vulgatus* (SGB1814) and *Bifidobacterium longum* (SGB17248) were detected in all westernized datasets as significantly transmitted between mothers and infants (Chi-squared tests, $P_{adj} < 0.05$; not prevalent enough in non-westernized datasets to assess transmissibility; Fig. 2c, Supplementary Table 16 and Methods). By contrast, other SGBs detected in infants—such as *Roseburia intestinalis* (SGB4951), which was found in 13 children and 102 mothers—were extremely rarely maternally transmitted (Supplementary

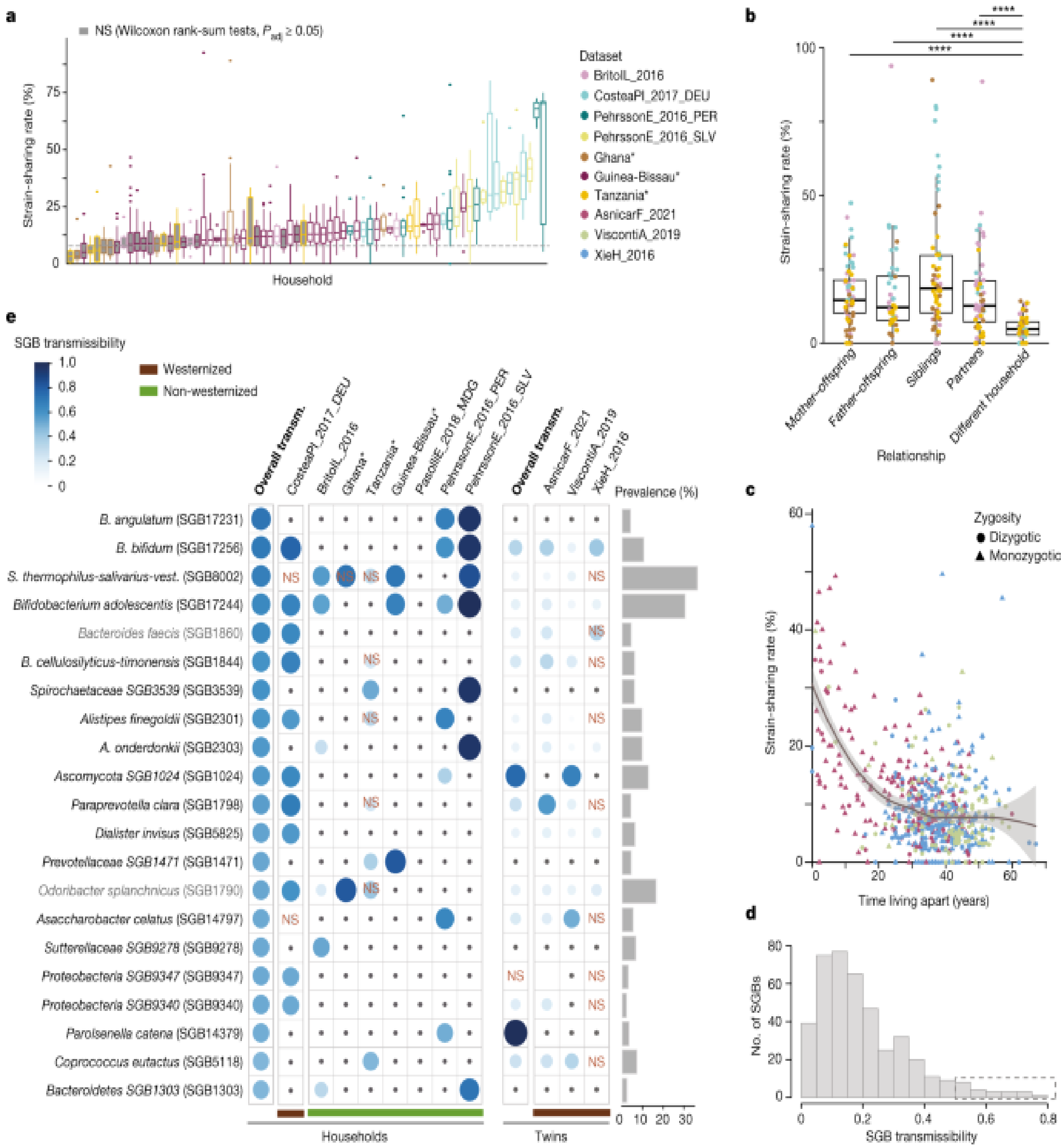
Table 9). The highly maternally transmitted SGBs were found to be gradually less shared between mothers and older offspring (Fig. 2c and Supplementary Table 16), but significant transmissibility of 52% of the highly maternally transmitted SGBs was detected even in senior individuals (50–85 years old) not cohabiting with their mothers (Fig. 2c and Supplementary Table 16).

Cohabitation drives transmission

Gut microbiome similarities among household members are well documented[45,47,48,49], but because of the missing strain-level resolution, most studies have not been able to conclude whether similarities at higher taxonomic levels reflected microorganism transmission or rather modulation by similar conditions (for example, genetics or diet). To examine horizontal gut microbiome transmission, we assessed strain sharing among 883 cohabiting individuals (up to 4 years old) in 212 households from 8 populations on 4 continents (Fig. 1a) with remarkably diverse lifestyles: from traditional subsistence in rural areas[17,23,30,34], to crowding conditions in large developing cities[23] and medium-sized industrialized affluent cities[27]. The majority of households displayed significantly higher person-to-person strain-sharing rates (between 11% and 71%) among cohabiting members than with non-cohabiting individuals of the same population (64% households, Wilcoxon rank-sum tests, $P_{adj} < 0.05$; 28% to 778% median increase in strain-sharing rates compared with among different households; Fig. 3a and Supplementary Table 17). Weaker differences were found for species-level microbiome similarities (beta diversity indices; Extended Data Fig. 4b) between individuals sharing households and non-cohabiting individuals (3% to 9% increase, Kruskal–Wallis tests with post hoc Dunn tests, $P_{adj} < 0.05$; Supplementary Table 8). Although person-to-person strain sharing varied largely across households (Kruskal–Wallis test, $n = 1,632$, $\chi^2 = 223$, $P = 2.8 \times 10^{-45}$), this was only slightly associated with westernized lifestyles (Wilcoxon rank-sum test, $n = 1,632$, $r = -0.22$, $P = 2.2 \times 10^{-18}$), possibly pointing to limited effects of environmental and social variables. Strain sharing between cohabiting individuals decreased with age (Wilcoxon rank-sum test for under 4 years of age versus 4 years and older, $n = 1,843$, $r = -0.12$, $P = 1.3 \times 10^{-7}$), supporting a lower colonization resistance in early

life[6,32]. By contrast, the number of strains of non-family origin (defined as those not shared with any household member) increased with age, as expected with increased cumulative exposure (Wilcoxon rank-sum test for under 4 years of age versus 4 years and older, $r = 0.20$, $P = 4.9 \times 10^{-8}$).

Fig. 3: Within-household and between-household gut microbiome transmission.



a, Pairwise person-to-person strain-sharing rates (number of shared strains/number of SGBs in common × 100%) in 72 households with at least

four cohabiting individuals ($n = 883$). The dashed line shows the median sharing rate among individuals in different households of the same village. Grey-filled boxes represent households with intra-household strain-sharing rates that are not significantly higher than inter-household sharing rates in the same population (Wilcoxon rank-sum two-sided tests, $P_{adj} \geq 0.05$; Supplementary Table 17). In box plots, box edges delineate lower and upper quartiles, the centre line represents the median and whiskers extend to 1.5 times the IQR. Novel datasets from the present study are highlighted with asterisks. **b**, Strain-sharing rates between individuals in households. Post hoc Dunn two-sided tests, $n = 282$, ****$P_{adj} < 10^{-4}$ (Supplementary Table 18). In box plots, box edges delineate lower and upper quartiles, the centre line represents the median and whiskers extend to 1.5 times the IQR. **c**, Strain-sharing rate in non-cohabiting adult twins ($n = 1{,}734$) decreases as a function of the time spent living apart (loess curve). The shaded area shows the 95% confidence interval. **d**, Histogram of household SGB transmissibility. **e**, A panel of 21 SGBs that are highly transmitted in households (SGB transmissibility >0.5 and significantly higher intra-household than inter-household transmissibility) showing their transmissibility in specific datasets and in non-cohabiting adult twins. NS, non-significant SGB transmissibility in the category (Chi-squared test on the number of transmitted and non-transmitted SGBs between household pairs and between pairs in different households; Supplementary Table 20). Only comparisons with at least three possible transmissions (species shared by at least three cohabiting pairs) are shown; comparisons with less than three possible transmissions are marked with a dot. Prevalence is defined as the percentage of samples in which the SGB was detected. Novel datasets from the present study are highlighted with asterisks. SGB names in grey use a strain identity threshold of 5% inter-individual nGD (Supplementary Table 4). *S. thermophilus-salivarius-vest.*, *S. thermophilus*, *S. salivarius* and *S. vestibulari*.

We next assessed strain sharing between parents and offspring, between siblings and between partners in the four populations in which kinship was known. All family relationships displayed significantly higher strain-sharing rates than different-household comparisons (post hoc Dunn tests, $n = 282$, $P_{adj} < 0.05$; Fig. 3b and Supplementary Table 18), but no significant differences were detected among them. Maternal and paternal strain-sharing rates were similar in children 4 years of age and older, and there was slightly

(but not significantly) higher strain sharing between younger (that is, less richly colonized), genetically related siblings than between partners. To assess the extent to which co-housing impacts strain sharing later in life, we analysed metagenomes from non-cohabiting adult twins who had lived together in the past (1,734 samples from three published cross-sectional datasets[2,25,33] in the United Kingdom), including both monozygotic and dizygotic twins. Strain sharing between twin pairs decreased significantly with the number of years spent living apart (Spearman's test, $n = 708$, $\rho = -0.30$, $P = 9.2 \times 10^{-15}$) and after accounting for their age (generalized linear model (GLM), $n = 648$, $\beta = -0.58$, $P = 7.1 \times 10^{-18}$; Fig. 3c). There was a moderate genetic effect beyond the influence of past cohabitation, with monozygotic twins displaying higher strain-sharing rates decades after cohabitation than dizygotic twins (Wilcoxon rank-sum tests, $P_{adj} < 0.05$; Extended Data Fig. 7 and Supplementary Table 19). Finally, the more gradual decline in age-associated strain sharing when partialling out the number of years twins have lived apart (GLM, $n = 648$, $\beta = -3.9 \times 10^{-3}$, $P = 0.02$) provides further evidence for the effect of cohabitation on microbiome transmission in adults and its larger quantitative effect than genetics and age. Strain sharing among adult twins might therefore be more the result of past cohabitation than of a long-lasting effect of shared transmission from their parents.

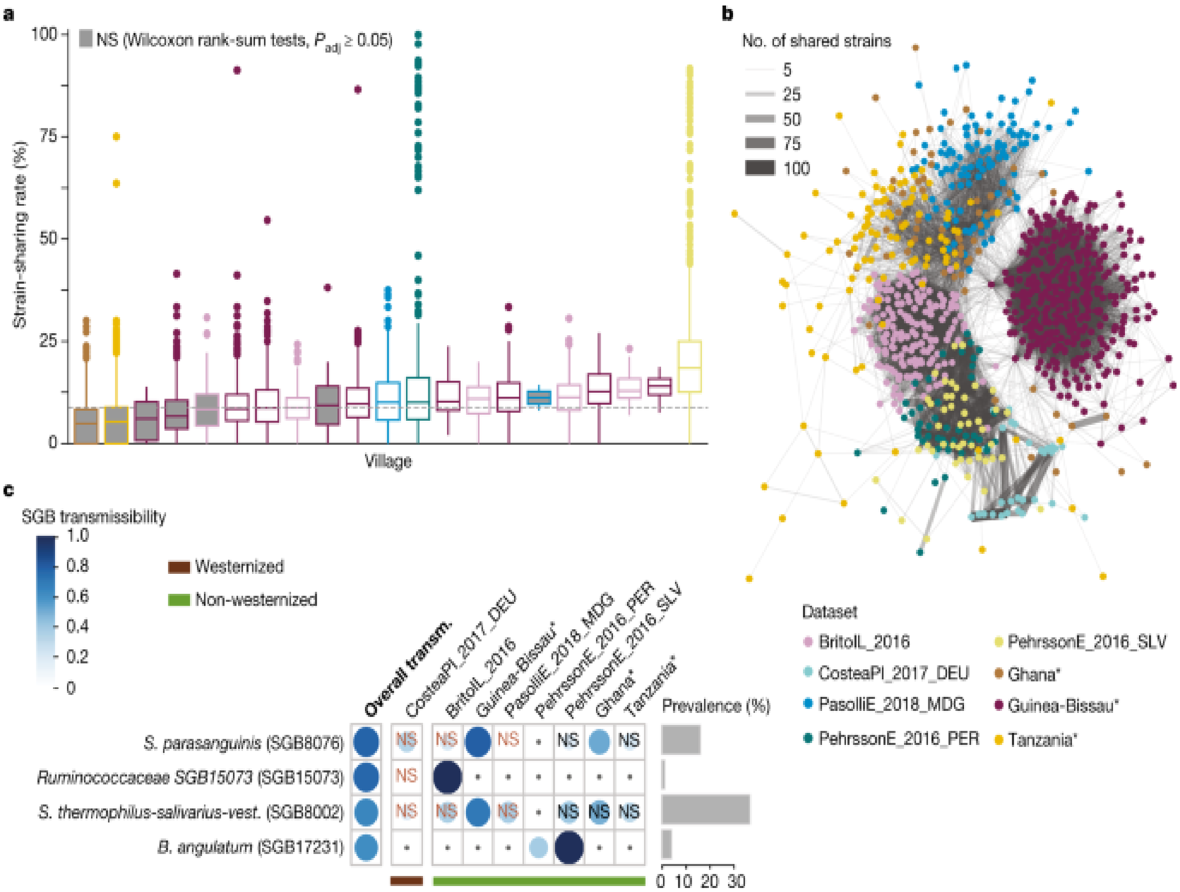
A panel of 21 SGBs (4% of assessed SGBs) from 10 different bacterial genera were highly transmitted between household members (SGB transmissibility >0.5 and significantly higher intra-household than inter-household transmissibility; Fig. 3d,e, Supplementary Table 20 and Methods). Household SGB transmissibility was not consistent across datasets (pairwise Spearman's tests, $P_{adj} \geq 0.05$; Supplementary Table 21), in contrast to mother-to-infant transmissibility, and we observed large differences in SGB transmissibility between westernized and non-westernized lifestyles (Fig. 3e) in concordance with their divergent microbiome composition[30,45,50,51]. A high portion (38%) of highly transmitted SGBs were species without characterized isolates or genomes (uSGBs) for the species ($n = 1$) or genus ($n = 7$) they belong to. Most highly transmitted *Bifidobacterium* and *Bacteroides* species in households coincided with those found highly transmitted from mother to offspring (Figs. 2c and 3e), suggesting these are efficient spreaders regardless of

transmission mode, in contrast to *Bifidobacterium angulatum* (SGB17231), which emerged as preferentially transmitted across households. Notably, SGBs that were highly transmitted within households tended to remain shared among twin pairs who moved apart (94% of the 21 highly transmissible SGBs; Fig. 3e and Supplementary Table 20), supporting the partial persistence of transmitted strains.

Microorganism transmission along populations

Non-cohabiting individuals in a village displayed non-negligible strain sharing of gut microbiome, in contrast to individuals with no presumed shared environments, albeit at notably lower rates than same-household members (Kruskal–Wallis test, $n = 1{,}132$ samples across 7 datasets, $\chi^2 = 1{,}721$, $P < 2.2 \times 10^{-16}$; post hoc Dunn tests, $P_{adj} < 0.05$; Extended Data Fig. 8a and Supplementary Table 22). Whereas intra-village strain-sharing rates were largely variable within populations (Fig. 4a), in 67% of villages, individuals from different households in the same villages had significantly higher strain-sharing rates than those in different villages (Wilcoxon rank-sum tests, $P_{adj} < 0.05$; Supplementary Table 23) in 5 out of the 7 populations assessed. Person-to-person microbiome transmission thus also occurs upon interaction between more distant contacts, and is potentially affected by population structures[4,17]. Indeed, we found that microbiome strain transmission within and between populations recapitulated host population structures (PERMANOVA on Euclidean distance in unsupervised strain-sharing network, $n = 951$, $R^2 = 46\%$, $P = 10^{-2}$; Fig. 4b and Methods) at a markedly stronger degree than that of species sharing (PERMANOVA on Euclidean distance on species sharing network, $n = 951$, $R^2 = 11\%$, $P = 10^{-2}$; Extended Data Fig. 8b).

Fig. 4: Gut microbiome transmission along villages and populations.



a, Person-to-person strain-sharing rates in different households of a village ($n = 1,132$). The dashed line shows the median strain-sharing rate among individuals in different villages of the same dataset. In box plots, box edges delineate lower and upper quartiles, the centre line represents the median and whiskers extend to 1.5 times the IQR. Grey-filled boxes show non-significant differences between the within village and inter-village person-to-person strain-sharing rate (Wilcoxon rank-sum two-sided tests, $P_{adj} \geq 0.05$; Supplementary Table 23). **b**, Gut microbiome strain-sharing unsupervised network of individuals in household datasets displaying population structure. Line width is proportional to the number of shared strains. **c**, Highly transmitted SGBs between individuals in different households (SGB transmissibility >0.5 and significantly higher intra-population than inter-population transmissibility), and transmissibility of these SGBs in specific datasets (classified by westernization status). NS, non-significant SGB transmissibility in the category (Chi-squared two-sided tests on the number of transmitted and non-transmitted SGBs between inter-household pairs and between pairs in different datasets; Supplementary

Table 24). Only comparisons with at least three possible transmissions (species shared by at least three pairs) are shown; comparisons with less than three possible transmissions appear with a dot. Prevalence is defined as the percentage of samples in which the SGB was detected. Novel datasets from the present study are highlighted with asterisks.

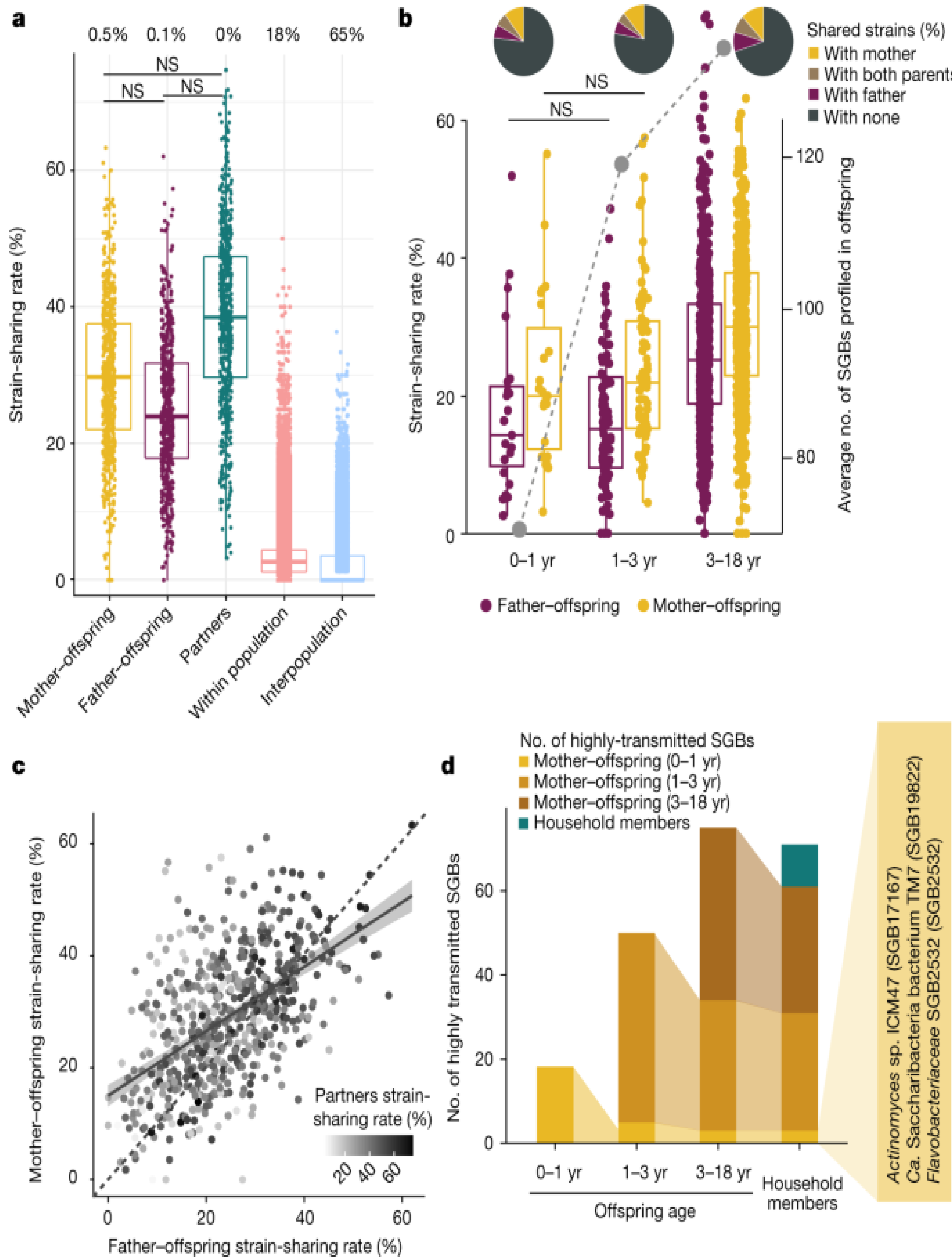
Although only 4 SGBs (0.8%) displayed high intra-population transmissibility overall (SGB transmissibility >0.5 and significantly higher intra- than inter-population transmissibility; Fig. 4c, Supplementary Table 24 and Methods), intra-population species transmissibility was highly consistent across datasets (pairwise Spearman's tests on SGB intra-population transmissibility by dataset, $\rho > 0$, $P_{adj} < 0.05$; Supplementary Table 25). Three highly transmitted SGBs are known members of the human microbiome: *B. angulatum* (SGB17231, 4% prevalence), *Streptococcus parasanguinis* (SGB8076, a species with opportunistic pathogen representatives[52], 16%), and *S. thermophilus*, *S. salivarius* and *S. vestibularis* (SGB8002, including some strains commonly used as probiotic[53], 37%), suggesting that both health-associated and potential pathogenic species can be efficient spreaders. A so-far uncharacterized species of the Ruminococcaceae family was also among the highly transmitted SGBs (SGB15073, 1% prevalence). Although *S. thermophilus*, *S. salivarius*, *S. vestibularis* and *B. angulatum* also appeared as highly transmitted in households, the specific high transmissibility of *S. parasanguinis* and SGB15073 among non-cohabiting individuals (Figs. 2c and 3e) suggests distinct spreading mechanisms.

Mostly horizontal oral transmission

Oral microbiome strains are probably more easily transmitted among individuals than gut strains, as saliva can be a direct vehicle[54], but person-to-person oral microbiome transmission remains underexplored[17,54,55]. We assessed the patterns of oral strain sharing in 1,929 newly sequenced metagenomes from households in the United States (USA dataset) together with 140 saliva metagenomes publicly available from a population in the Fiji islands[17] by strain-level profiling of 252 SGBs (Methods). We detected a strain-sharing rate gradient across shared environments and kinship, similar

to that observed for gut microbiome strain sharing: cohabiting individuals displayed 32% median oral strain-sharing rates, whereas non-cohabiting individuals in the same or different populations shared 3% and 0%, respectively (Kruskal–Wallis test, $N = 2{,}069$, $\chi^2 = 41{,}317$, $P < 2.2 \times 10^{-16}$; Fig. 5a). Cohabiting individuals thus feature 10 times higher oral strain-sharing rates than non-cohabiting individuals in the same population, in contrast to less than 0.5 times higher species-level microbiome similarity (Extended Data Fig. 5b and Supplementary Table 26), suggesting that strain transmission between household members is a stronger driver of genetic microbiome composition than species-level microbiome convergence through similar conditions and lifestyles. In addition, less than 0.5% of same-household members did not share a single strain, in contrast to 18% of intra-population pairs and 65% of inter-population pairs; this indicates that person-to-person transmission of bacterial oral strains occurs more frequently than gut microbiome transmission (Fig. 1f).

Fig. 5: Transmission of the oral microbiome.



a, Person-to-person strain-sharing rates (number of shared strains/number of SGBs in common × 100%) across relationships ($n = 2{,}069$). All comparisons are statistically significant unless stated otherwise (Kruskal–Wallis test, $n =$

26,218, $\chi^2 = 11{,}420$, $P < 2.2 \times 10^{-16}$, post hoc Dunn two-sided tests, $P_{adj} < 0.05$; Supplementary Table 28). Numbers along the top show the percentage of pairs between which no strain-sharing event was detected. **b**, Mother–offspring and father–offspring sharing rates (number of shared strains/number of SGBs in common × 100%) ($n = 2{,}069$) (left axis; box plot and non-grey dots) and median number of SGBs detected in the offspring (right axis; grey dots). Post hoc Dunn two-sided tests, Supplementary Table 29. All comparisons are statistically significant after correction for multiple testing unless stated otherwise. In box plots, box edges delineate lower and upper quartiles, the centre line represents the median and whiskers extend to 1.5 times the IQR. Pie charts show the percentage of strains shared between pairs of individuals. **c**, Strain sharing across cohabiting individual relationships are positively correlated (Spearman's two-sided tests, mother–offspring and father–offspring: $n = 637$, $\rho = 0.52$, $P < 2.2 \times 10^{-16}$; mother–offspring and partners: $n = 611$, $\rho = 0.21$, $P = 1.2 \times 10^{-7}$; father–offspring and partners: $n = 611$, $\rho = 0.38$, $P < 2.2 \times 10^{-16}$). Dashed line is the diagonal, where mother–offspring strain-sharing rate is equal to father–offspring strain-sharing rate. The shaded area shows the 95% confidence interval. **d**, The persistence of highly transmitted SGBs (SGB transmissibility >0.5 and significantly higher intra-household than inter-household transmissibility) between mother and offspring across age categories and among household members who are at least four years of age. *Ca.*, *Candidatus*.

Distinct age- and kinship-associated patterns emerged: in contrast to the gut microbiome pattern, oral strain-sharing rates increased with offspring age (Spearman's test, $n = 658$, $\rho = 0.15$, $P = 1.9 \times 10^{-4}$ for mother–offspring and $n = 643$, $\rho = 0.24$, $P = 7.1 \times 10^{-10}$ for father–offspring), especially after 3 years of age (Kruskal–Wallis test, $\chi^2 = 31$, $P = 1.7 \times 10^{-7}$ for mother–offspring, $\chi^2 = 58$, $P = 2.4 \times 10^{-13}$ for father–offspring, post hoc Dunn tests, Supplementary Table 27), coinciding with the increasing accumulation of microorganism species in the offspring's oral microbiome (from a median of 49 shared species between mothers and offspring and 55 shared species between fathers and offspring up to 1 year of age, to a median of 85 shared species between mothers and offspring and 86 shared species between fathers and offspring up to 18 years of age; Spearman's test, $n = 658$, $\rho = 0.21$, $P = 6.2 \times 10^{-8}$; Fig. 5b). No significant differences were detected

among different types of relationships (post hoc Dunn tests, $P_{adj} \geq 0.05$; Supplementary Table 28), but strain-sharing rates were slightly higher between partners (median 38%) than for the younger offspring with their mothers (30%) and fathers (24%; Fig. 5a) probably reflecting greater intimacy[54]. Mother–offspring species sharing rates tended to be higher than father–offspring species sharing rates across age ranges (post hoc Dunn tests, $P_{adj} < 0.05$; Supplementary Table 29), potentially as a result of closer contacts and imprinting through breastfeeding. However, although the proportion of strains shared with both partners increased slightly with offspring age (6% below 1 year to 8% below 18; Fig. 5b), even more strains were shared with each parent separately (17–21% with mothers and 13–17% with fathers). Overall, parental strain transmission does not seem to particularly seed oral microbiome assembly in early life, but rather appears to exploit horizontal transmission modes that are also dependent on the duration of the contact.

Intra-family oral strain transmission varied largely across households (0–75%), and although conclusions on lifestyle associations cannot be drawn on the basis of the two datasets available with disparate sample sizes, we did find significant correlations between strain sharing in households across all types of kinship assessed (Fig. 5c). Mother–offspring strain-sharing rates correlated with father–offspring strain-sharing rates (Spearman's test, $n = 637$, $\rho = 0.52$, $P < 2.2 \times 10^{-16}$) and with partner strain-sharing rates (Spearman's test, $n = 611$, $\rho = 0.21$, $P = 1.2 \times 10^{-7}$). Also, father–offspring strain-sharing rates correlated with those between partners (Spearman's test, $n = 611$, $\rho = 0.38$, $P < 2.2 \times 10^{-16}$). Closely interacting households thus seem to favour oral strain transmission among all cohabiting individuals regardless of kinship.

We next assessed parent-to-offspring and household oral species transmissibility (Supplementary Table 30). Eighteen SGBs (half of which were uSGBs) from 16 different genera were significantly highly shared between mothers and their infants up to 1 year of age (19% of total SGBs assessed, SGB transmissibility>0.5 and significantly higher intra-mother–offspring pair than inter-mother–offspring pair transmissibility; Fig. 5d), including two *Prevotella* species (*Prevotella histicola* (SGB1543) and *Prevotella pallens* (SGB1564)) and two largely uncharacterized *Actinomyces*

species (SGB17132 and SGB17167; Supplementary Table 31). Although SGB transmissibility up to 1 year of age showed a strong correlation with that at 1–3 years of age (Spearman's test, $n = 95$, $\rho = 0.73$, $P < 2.2 \times 10^{-16}$) and between 3 and 18 years of age ($n = 95$, $\rho = 0.78$, $P < 2.2 \times 10^{-16}$), only five species persisted as highly transmitted between the first (up to 1 year) and second (1 to three years) age bin and three persisted to the third (3 to 18 years) age bin, with up to 68 further species appearing (Fig. 5d and Supplementary Table 31). These 68 later-emerging species were highly concordant with the 70 species (including 28 uSGBs) displaying significantly high household transmissibility (28% of total SGBs assessed; Supplementary Table 32), including the three persisting highly maternally transmitted SGBs. By contrast, no species was highly transmitted among non-cohabiting individuals (Supplementary Table 30). Overall, three under-characterized SGBs thus exhibited consistently strong oral transmission potential: *Actinomyces* sp. ICM47 (SGB17167), *Candidatus* Saccharibacteria bacterium TM7 (SGB19822), and a uSGB of the family Flavobacteriaceae (SGB2532) (Fig. 5d and Extended Data Fig. 9).

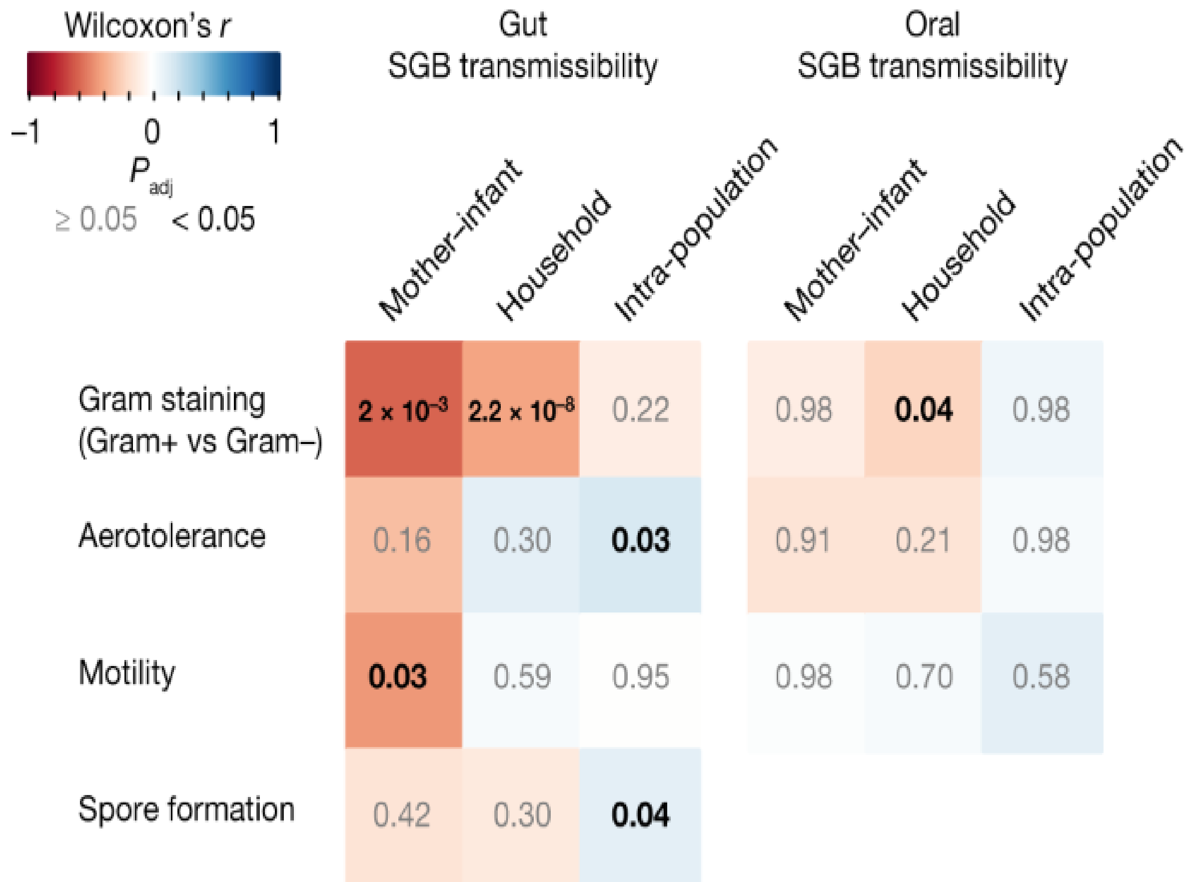
Phenotypes linked to transmission modes

The transmissibility of gut species was highly consistent across geographically distant datasets with diverse lifestyles (Spearman's tests, $P_{adj} < 0.05$; mother-to-infant: 71%, intra-population: 75% significant associations; Supplementary Tables 15, 21 and 25, with transmissibility estimates ranging between 0 and 100%). At the same time, gut species were often preferentially transmitted through specific modes[56] (23% SGBs were highly transmitted through more than 1 mode; Figs. 2c, 3e and 4c). By contrast, highly transmitted oral SGBs across transmission modes were largely overlapping (Fig. 5d). Species transmissibility did not seem to predominantly follow a mass-action model of transmission—neither median relative abundance nor the prevalence of a species in populations was positively associated with its transmissibility (Spearman's one-sided tests, $P_{adj} \geq 0.05$; Supplementary Table 33).

The absence of a direct link between prevalence and transmissibility is consistent with species transmissibility through different modes being a

specific trait, so we next explored whether phenotypic properties associated with persistence in the environment[3,4] could better account for the patterns we detected. As 58% of the gut and 24% of the oral SGBs that we profiled at the strain level have not yet been cultured, we inferred bacterial phenotypes on the basis of their genome sequences (Methods). The predicted phenotypes showed more than 90% concordance with experimentally determined traits in cases where those were available (Supplementary Table 34 and Methods). Gut and oral microbiome transmission modes were associated with specific phenotypic properties (Fig. 6). Gram-negative bacteria—generally more resistant to sanitizers and disinfectants[57]—displayed enhanced gut maternal and household transmissibility (Wilcoxon rank-sum tests on first versus fourth quartiles of SGB transmissibility, $n = 35$, $r = -0.59$, $P_{adj} = 2.0 \times 10^{-3}$ and $n = 213$, $r = -0.40$, $P_{adj} = 2.2 \times 10^{-8}$, respectively), together with increased oral household transmissibility ($n = 126$, $r = -0.22$, $P_{adj} = 0.04$). Longer-range gut intra-population transmissibility required more powerful environmental survival mechanisms—that is, aerotolerance and spore formation ($n = 268$, $r = 0.16$, $P_{adj} = 0.03$ and $n = 280$, $r = 0.10$, $P_{adj} = 0.04$, respectively). With less than 10% of profiled gut SGBs being predicted as oxygen-resistant in contrast to more than 66% of oral ones, aerotolerance was not associated with transmissibility of oral SGBs (Fig. 6). Finally, the motile species that are frequent but unstable inhabitants of the infant gut[58] were less frequently transmitted from mothers to offspring than non-motile SGBs ($n = 35$, $r = -0.43$, $P_{adj} = 0.03$), which could be beneficial given the link between motility and virulence[59]. Overall, our results suggest that microorganism phenotypic properties promoting survival in the environment at least partially modulate person-to-person gut microbiome transmission dynamics, whereas a notably weaker link was found for oral microbiome transmission.

Fig. 6: Association between gut and oral species transmissibility and phenotypical properties.



SGB phenotypes were inferred using Traitar[60] (Methods). Association between SGB transmissibility and predicted phenotypes was assessed with Wilcoxon rank-sum two-sided tests on the 25% of SGBs displaying the highest transmissibility and compared with the 25% of SGBs displaying the lowest transmissibility for each transmission mode and environment. Colours represent the Wilcoxon *r* statistics; significant P_{adj} values are shown in black ($P_{adj} < 0.05$) and in grey otherwise.

Conclusion

Our integrative multi-cohort study of microbiome transmission across diverse populations shows extensive previously overlooked person-to-person transmission. This corroborates already suggested hypotheses[3,4,5,16] and reveals that the transfer of microorganism strains among individuals in long-lasting close contact is a major driver in shaping the personal genetic makeup of the microbiome, and thus of the corresponding metabolic and

host–microorganism interaction potential. Although strain sharing was, as expected, greatest between mother and infant gut microbiomes during the first year of life[9,10,12,29,32] (median of 50%), shared strains also accounted for 12% and 32% of the gut and oral microbiome species in common between cohabiting individuals, respectively (Figs. 1f and 5a). Such an effect might be induced by close physical interaction even when such interaction started only in adulthood (13% and 38% gut and oral strain sharing between partners respectively; Figs. 3b and 5a) and is partially reversible over long periods, with twins decreasing their initial strain sharing of around 30% to about 10% over 30 years of living apart (Fig. 3c). Because unrelated individuals in different populations or even in different villages of the same population share hardly any strains (0% median strain-sharing rate), our results highlight a non-negligible effect of social interactions in shaping the microbiome, which could have a role in microbiome-associated diseases, and warrants consideration of person-to-person strain transmission in human microbiome studies.

By contrast, we found little influence of divergent lifestyles on microbiome transmission dynamics: despite massive microbiome composition differences in populations loosely defined as westernized or non-westernized[34,43,51] on the basis of characteristics such as diet, access to medical facilities and drugs, and hygiene conditions (Methods), we found remarkably similar vertical and horizontal strain-sharing rates. Larger, diverse cohorts and more detailed metadata on participants' lifestyles and cultural practices are needed to ensure the robustness of this finding, but our results might point to similar microorganism colonization resistance in different populations that could be of greater importance in establishing durable colonization than the intrinsic rates of transmission events. Our results also suggest that the higher richness of microorganisms observed in non-westernized communities[34,43] is not promoted by enhanced transmission from other household members, but is rather a consequence of the interaction with the environment as well as diets and lifestyles supporting microorganism diversity.

Species showing particularly high transmissibility (Figs. 2c, 3e, 4c and 5d) should be the starting point for a deeper understanding of the genomic and phenotypic characteristics that can in turn inform transmission mechanisms.

Although our study could not resolve whether person-to-person microbiome transmission was direct or its directionality, it provided a systematic overview of microbiome transmission in humans. Further insight into person-to-person microbiome transmission and its directionality could be obtained using specific study designs modelling changes in routine social-interaction networks in humans (for example, following household changes) or in other social animals. The improved strain tracking methods we used that included strain-level profiling of so-far uncultured species[39] and species-specific definitions of strain based on phylogenetic distances enabled us to scale to large numbers of samples corresponding to more than 800,000 strains. Nonetheless, future studies with whole-genome resolution enabled by deeper sequencing, long-read technologies or single-cell approaches may enable further clarification and refinement of these findings. Overall, our results reinforce the hypothesis that several diseases and conditions that are currently considered non-communicable should be re-evaluated[5], and that accounting for transmissibility and social network structure will improve the design of future microbiome investigations and modulation approaches.

Methods

Metagenomic datasets

A total of 9,715 samples from 31 human metagenomic datasets (total: 5.17×10^{11} reads, average: 5.32×10^{7} reads per sample) with available metadata to enable assessment of microbiome transmission between healthy mothers and offspring, households, twin pairs, villages and populations (that is, cohabitation information) were selected for inclusion in this study (Supplementary Tables 1 and 2). We also included publicly available stool shotgun metagenomic datasets with samples from at least 15 healthy individuals to whom no intervention (such as antibiotic or drug treatment, or specific diet) was performed, with at least 2 of the samples taken less than 6 months apart to assess within-subject strain retention and set species-specific operational definitions of strain identity 25 datasets were publicly available, three of which were expanded in this study with 14 (FerrettiP_2018[9]), 32 (Ghana dataset[34]) and 61 (Tanzania dataset[34]) samples. Newly included

samples were collected and processed following the protocols described in the original publications. In addition, eight datasets (total: 2,800 samples) were newly collected and sequenced in the context of this study as described below, using similar methods (although differences in sample processing, DNA extraction and sequencing library preparation do not directly affect the phylogenetic distances that we use to infer strain sharing).

Consistent metadata collection and organization

We retrieved the metadata on sample and subject identifiers, time points, participant's age, gender, mode of delivery (vaginal or caesarian section), family identifiers, family relationships, twin zygosity and age at which twins moved apart, village, and country from curatedMetagenomicData 3.0.0 (ref. [61]) when included in the resource, and from the publications' supplementary materials or specified repository otherwise. Metadata of all metagenomes, including newly sequenced samples, were curated and organized in the curatedMetagenomicData format and are available in Supplementary Table 2. Partners were defined as couples that share a household. Populations were classified on the basis of their westernization status (westernized or non-westernized), considered as the adoption of a westernized lifestyle and not in geographical terms, and defined as intake of diets typically rich in highly processed foods (with high fat content, low in complex carbohydrates and rich in refined sugars and salt), access to healthcare and pharmaceutical products, hygiene and sanitation conditions, reduced exposure to livestock, and increased population density. The classification was based on the information available on how populations included in the study differ on the above criteria and how the samples were reported in the original publications. While we acknowledge that this binary classification has evident limitations[62], it enables insight into the association of person-to-person microbiome transmission with host lifestyle.

Newly sequenced metagenomic datasets

Argentina dataset

A total of 14 mothers (16–37 years old) and 13 of their infants below 1 year of age in rural areas in Argentina (villages of Villa Minetti, Esteban Rams, Pozo Borrado, Las Arenas, Cuatro Bocas, Logroño, Montefiore and Belgrano; Santa Fe province; Supplementary Table 2)—considered here as a non-westernized population—were enroled in the study. DNA was extracted from faecal samples using the QIAamp DNA stool kit (Qiagen) following the manufacturer's instructions. Sequencing libraries were prepared using the Nextera DNA Flex Library Preparation Kit (Illumina), following the manufacturer's guidelines. Sequencing was performed on the Illumina NovaSeq 6000 platform following manufacturer's protocols.

Colombia dataset

A total of 12 mothers (15–40 years old) and 12 of their infants below 6 months of age from communities of the Wayúu ethnic group from the Caribbean Region in Colombia (communities of Etkishimana, Koustshachon, Paraiso, Invasión, Tocomana, Warruptamana and Wayawikat; Supplementary Table 2)—considered here as a non-westernized population—were enroled in the study. DNA from stool samples was extracted using the Master-Pure DNA extraction Kit (Epicentre) following the manufacturer's instructions with the following modifications: samples were treated with lysozyme (20 mg ml^{-1}) and mutanolysin (5 U ml^{-1}) for 60 min at 37 °C and a preliminary step of cell disruption with 3-μm diameter glass beads during 1 min at 6 m s^{-1} by a bead beater FastPrep 24-5G Homogenizer (MP Biomedicals). Purification of the DNA was performed using DNA Purification Kit (Macherey–Nagel) according to manufacturer's instructions. DNA concentration was measured using Qubit 2.0 Fluorometer (Life Technologies) for further analysis. Sequencing libraries were prepared using the Nextera DNA Flex Library Preparation Kit (Illumina), following the manufacturer's guidelines. Sequencing was performed on the Illumina NovaSeq 6000 platform following manufacturer's protocols.

China_1 dataset

A total of 116 nonagenarians and centenarians (97 female, 19 male, 94–105 years old) and 231 of their offspring (79 female, 152 male, 50–85 years old) in the city of Qidong (Jiangsu province, China) were enroled (considered

here as a westernized population)[63]. All participants were free of major illnesses at the time of inclusion. Fresh stool samples were collected at the Shanghai Tenth Hospital, and stored at −20 °C upon collection. DNA was extracted using the EZNA Stool DNA Kit (Omega Bio-tek) following manufacturer's instructions. DNA integrity and size were evaluated by 1% agarose gel electrophoresis, and DNA concentrations determined with NanoDrop (Thermo Fisher Scientific). DNA libraries were constructed according to the TruSeq DNA Sample Prep v2 Guide (Illumina), with 2 µg of genomic DNA and an average insert size of 500 bp. Library quality was evaluated with a DNA LabChip 1000 Kit (Agilent Technologies). Sequencing was conducted on an Illumina HiSeq 4000 platform with a 150 bp paired-end read length.

China_2 dataset

A total of 8 mothers and 19 infants below 1 year of age in a rural population in China (Bin county, Shaanxi province, northwest China) were enroled as part of a larger study (ClinicalTrials.gov NCT02537392); they were considered here as a non-westernized population. DNA was extracted with the QIAamp Fast DNA Stool Mini Kit (Qiagen), and precipitated with ethanol. Sequencing libraries were prepared using the Nextera DNA Flex Library Preparation Kit (Illumina), following the manufacturer's guidelines. Sequencing was performed on the Illumina NovaSeq 6000 platform following manufacturer's protocols.

Guinea-Bissau dataset

Samples from 342 volunteers (0–85 years old) in 74 households in the island of Bubaque (Bijagos Archipelago, Guinea-Bissau)—considered here as a non-westernized population—were collected and DNA extracted as part of a previous study[64]. In brief, samples were frozen at −20 °C at a reference laboratory. After homogenization and washing, DNA was extracted using the DNeasy PowerSoil PRO kit (Qiagen) with custom modifications[64]. Sequencing libraries were prepared using the Nextera DNA Flex Library Preparation Kit (Illumina), following the manufacturer's guidelines. Sequencing was performed on the Illumina NovaSeq 6000 platform following manufacturer's protocols.

Italy_1 dataset

A total of 4 mothers (37–46 years old) and their 8 children (0–2 years old) were enroled at the Santa Chiara Hospital in Trento, Italy; they were considered here as a westernized population. Mother stool samples were collected during or shortly after the delivery by the hospital staff, using faecal material collection tubes (Sarstedt). Infant stool samples were collected by the mothers, frozen at −20 °C upon collection and moved to a −80 °C facility within a week. 48 samples were collected in total (Supplementary Table 2). DNA was extracted using the PowerSoil DNA Isolation Kit (MoBio Laboratories), as described in the HMP protocol (Human Microbiome Project Consortium)[65], with addition of a preliminary heating step (65 °C for 10 min, 95 °C for 10 min). DNA was recovered in 10 mM Tris pH 7.4 and quantified using the Qubit 2.0 (Thermo Fisher Scientific) fluorometer per the manufacturer's instructions. Sequencing libraries were prepared using the NexteraXT DNA Library Preparation Kit (Illumina), following the manufacturer's guidelines. Sequencing was performed on the Illumina HiSeq 2500 platform.

Italy_2 dataset

A total of 19 mothers (30–47 years old) and 37 healthy children (0–11 years old) were enroled at the IRCCS Istituto Giannina Gaslini in Genoa, Italy as part of a larger study, considered here as a westernized population. Stool samples were collected in DNA/RNA shield faecal collection tubes (Zymoresearch) and stored at −80 °C until DNA extraction. DNA extraction was performed with the DNeasy PowerSoil Pro Kit (Qiagen) according to the manufacturer's procedures. DNA concentration was measured using the NanoDrop spectrophotometer (Thermo Fisher scientific) and stored at −20 °C. Sequencing libraries were prepared using the NexteraXT DNA Library Preparation Kit (Illumina), following the manufacturer's guidelines. Sequencing was performed on the Illumina NovaSeq 6000 platform following manufacturer's protocols.

USA dataset

A total of 1,929 saliva samples from 646 families in the NY Genome Center Cohort of the SPARK collection (Western IRB (https://www.wcgirb.com/), protocol tracking number: WIRB20151664, considered here as a westernized population) were included in the analysis, consisting of 640 mother samples (22–55 years old), 631 father samples (23–67 years old), and 658 samples from normally developing offspring (0–18 years old). Saliva was collected using the OGD-500 kit (DNA Genotek), and DNA was extracted using a Chemomagic MSM1/360 DNA extraction instrument and eluted into 110ul of TE buffer at PreventionGenetics (Marshfield). Sequencing libraries were prepared with the Illumina DNA PCR-Free Library Prep kit (Illumina), following the manufacturer's guidelines. Sequencing was performed on the Illumina NovaSeq 6000 platform using S2/S4 flow cells and following manufacturer protocols.

Metagenome pre-processing and quality control

Newly sequenced stool samples were pre-processed using the pipeline described at https://github.com/SegataLab/preprocessing. Shortly, metagenomic reads were quality-controlled and reads of low quality (quality score <Q20), fragmented short reads (<75 bp), and reads with >2 ambiguous nucleotides were removed with Trim Galore (v0.6.6). Contaminant and host DNA was identified with Bowtie2 (v2.3.4.3)[66] using the -sensitive-local parameter, allowing confident removal of the phiX 174 Illumina spike-in and human-associated reads (hg19 human genome release). Remaining high-quality reads were sorted and split to create standard forward, reverse and unpaired reads output files for each metagenome.

Newly sequenced saliva samples were pre-processed using a custom version of the pipeline described in https://github.com/SegataLab/preprocessing. Shortly, metagenomic reads were quality-controlled, removing reads of low quality (quality score <Q20), fragmented short reads (<75 bp), and reads with >2 ambiguous nucleotides. Contaminant and host DNA was identified with Bowtie2 (v2.3.5.1)[66] in 'end-to-end' global mode, allowing confident removal of human-associated reads (hg19). Remaining high-quality reads were sorted and split to create standard forward, reverse and unpaired reads output files for each metagenome.

Read statistics of stool and saliva samples (number of reads, number of bases, minimum and median read length per sample) are detailed in Supplementary Table 2. Metagenomes with ≥3 million reads were included in the analysis ($n = 7,646$ stool, $n = 2,069$ oral), while metagenomes with insufficient sequencing depth were excluded ($n = 97$ stool, $n = 0$ oral).

Expanded SGB database

A custom database containing 160,267 MAGs and 75,446 isolate sequencing genomes was retrieved from ref. [30], and expanded with 184 MAGs from the Italian mother–infant dataset[9] expanded in the current study, 1,439 MAGs from Italian centenarians[67], 3,584 MAGs obtained from stool samples of individuals in non-westernized populations[34], 2,985 MAGs from stool samples of non-human primates[68], 20,404 MAGs from cow rumen[69], 14,097 MAGs from mouse samples[70,71,72,73,74,75,76,77,78,79,80,81,82,83], 1,235 MAGs from termites (PRJNA365052, PRJNA365053, PRJNA365054, PRJNA365049, PRJNA365050, PRJNA365051, PRJNA405700, PRJNA405701, PRJNA405702, PRJNA405782, PRJNA405783, PRJNA366373, PRJNA366374, PRJNA366375, PRJNA366251, PRJNA405703, PRJNA366252, PRJNA366766, PRJNA366357, PRJNA366358, PRJNA366361, PRJNA366362, PRJNA366363, PRJNA366255, PRJNA366256, PRJNA366257, PRJNA366253, PRJNA405704, PRJNA366254 and PRJNA405781), 7,760 MAGs available from a previous catalogue[84], 2,137 MAGs from NCBI GenBank, and 63,142 reference genomes from NCBI GenBank (see https://github.com/SegataLab/MetaRefSGB for details). MAGs from the Italian mother–infant dataset, and those of non-human hosts were assembled using MEGAHIT[85], while those of the Italian centenarian dataset and non-westernized populations were assembled with metaSPAdes[86], using default parameters in both cases.

For the newly added MAGs we employed the following protocol on the metagenomic assemblies. Assembled contigs longer than 1,500 nucleotides were binned into MAGs using MetaBAT2[87]. Quality control of all genomes was performed with CheckM version 1.1.3 (ref. [88]), and only medium- and high-quality genomes (completeness ≥50% and contamination ≤5%) were

included in the database. Prokka version 1.12 and 1.13 (ref. [89]) were used to annotate open reading frames of the genomes. Coding sequences were then assigned to a UniRef90 cluster[90] by performing a Diamond search (version 0.9.24)[91] of the coding sequences against the UniRef90 database (version 201906) and assigning a UniRef90 ID if the mean sequence identity to the centroid sequence was above 90% and covered more than 80% of the centroid sequence. Protein sequences that could not be assigned to any UniRef90 cluster were de novo clustered using MMseqs2[92] within SGBs following the Uniclust90 criteria[93].

Genomes were clustered into species-level genome bins (SGBs) spanning ≤5% genetic diversity, and those to genus-level genome bins (GGBs, 15% distance) and family-level genome bins (FGBs, 30% distance), as described in ref. [30]. MAGs were assigned to SGBs by applying 'phylophlan_metagenomic', a subroutine of PhyloPhlAn 3 (ref. [94]), which uses Mash[95] to compute the whole-genome average nucleotide identity among genomes. When no SGB was below 5% genetic distance to a genome, new SGBs were defined, based on the average linkage assignment and hierarchical clustering (allowing a 5% genetic distance among genomes in the dendrogram). The same procedure was followed to assign SGBs to novel GGBs and FGBs when those were not yet defined.

Taxonomic assignment of SGBs and definition of kSGBs and uSGBs

SGBs containing at least one reference genome (kSGBs) were assigned the taxonomy of the reference genomes following a majority rule, up to the species level. SGBs with no reference genomes (uSGBs) were assigned the taxonomy of its corresponding GGB (up to the genus level) if this contained reference genomes, and of its corresponding FGB (up to the family level) if the latter contained reference genomes. If no reference genomes were present in the FGB, a phylum was assigned based on the majority rule applied on up to 100 closest reference genomes to the MAGs in the SGB as provided by 'phylophlan_metagenomic'. Taxonomic assignment of SGBs profiled at strain level in this study can be found in Supplementary Tables 3 and 4.

Species-level profiling of metagenomic samples

Species-level profiling was performed on all the 9,715 samples with MetaPhlAn 4 (refs. [38,39]) with default parameters and the custom SGB database. uSGBs with less than 5 MAGs were discarded as potential assembly artefacts or chimeric sequences and unlikely to reach the prevalence thresholds in the profiling. SGB core genes were defined as open reading frames in an existing UniRef90 or in a de novo clustered gene family (following the Uniclust90 clustering procedure[93]) present in at least half of the genomes (that is, 'coreness' 50%) of the SGB. Core genes were further optimized by selecting the highest coreness threshold that allowed retrieval of at least 800 core genes. Core genes of each SGBs were then screened to identify marker genes by checking their presence in other SGBs. This was done by a procedure that first divided core genes into fragments of 150 nt and then aligned the fragments against the genomes of all SGBs using Bowtie2 (version 2.3.5.1; -sensitive option)[66]. Marker genes were defined as core genes with no fragments found in at least 99% of the genomes of any other SGB. For SGBs with less than 10 marker genes, conflicts were defined as occurrences of more than 200 core genes of an SGB in more than 1% of genomes of another SGB, and conflict graphs were generated by retrieving all conflicts for that SGB. Each conflict graph was processed iteratively, retrieving all the possible merging scenarios, in order to get the optimal merges for the conflict that both minimize the number of merged SGBs and maximize the number of markers retrieved. Finally, for each SGB, a maximum of 200 marker genes were selected based first on their uniqueness and then on their size (bigger first), and SGBs still with less than 10 markers were discarded. Merged gut and oral SGBs (SGB_group) can be found in Supplementary Tables 3 and 4, respectively. The resulting 3.3M marker genes (189 ± 34marker genes per SGB(mean ± s.d.)) were used as a new reference database for MetaPhlAn and StrainPhlAn profiling.

Strain-level profiling of metagenomic samples

Strain profiling was performed with StrainPhlAn4[38,39] using the custom SGB marker database, with parameters "marker_in_n_samples 1 -sample_with_n_markers 10 –phylophlan_mode accurate -mutation_rates". To reduce noise, only SGBs detected in ≥20 samples and at least 10% of

samples in a dataset with ≥10 markers (-print_clades_only argument in StrainPhlAn) were selected for strain-level profiling (*n* = 646 and *n* = 252 SGBs in stool and oral samples respectively). The total of 200 marker genes was available for the majority of SGBs (*n* = 481/646 gut SGBs and *n* = 148/252 oral SGBs). The average coverage across SGBs was 1.3×. For the SGBs potentially derived from fermented foods, sequences of MAGs assembled in ref. [40] were added using parameter "-r". Compared to an assembly based approach (high-quality MAGs defined as >90% completeness and <5% contamination; assembly method reported in the section "Expanded SGB database" above), strain-level profiling with StrainPhlAn allowed strain-sharing assessment among species in many more samples (median of 355 strain-level profiles per SGB and interquartile range (IQR) = [185, 806] versus median of 69 high-quality MAGs per SGB and IQR = [7, 60]).

Detection of strain-sharing events

To detect strain-sharing events, we first set SGB-specific normalized phylogenetic distance (nGD) thresholds that optimally separated same-individual longitudinal strain retention (same strain) from unrelated-individual (different strain) nGD distributions in five published stool metagenomic datasets from four different countries (Germany, Kazakhstan, Spain and United States) on three continents[20,22,27,28,31]. nGDs were calculated as leaf-to-leaf branch lengths normalized by total tree branch length in phylogenetic trees produced by StrainPhlAn, which are built on marker gene alignments on positions with at least 1% variability. For SGBs detected in at least 50 pairs of same-individual stool samples obtained no more than 6 months apart (*n* = 145 SGBs; the two samples for a certain individual in which the species could be profiled at the strain level and that were closest in time were selected), nGD thresholds were defined based on maximizing Youden's index, and limiting at 5% the fraction of unrelated individuals to share the same strain as a bound on a false discovery rate (Extended Data Fig. 3). The assumption of frequent strain persistence in an individual for at least 6 months is supported by the distribution of phylogenetic distances in the longitudinal sets: for all species this has a peak at nGD approaching 0 (Extended Data Fig. 3), notably higher than that observed for inter-individual sample comparisons. For SGBs detected in less

than 50 same-individual close pairs ($n = 501$) and in oral samples ($n = 252$), for which species-specific nGD cannot be reliably estimated, the nGD corresponding to the 3rd percentile of the unrelated individual nGD distribution was used. This value is the median percentile of the inter-individual nGD distribution corresponding to the nGD maximizing the Youden's index of SGBs with at least 50 same-individual comparisons. The three sets of thresholds are thus three technical definitions of the same principle—that is, the individual specificity and the persistence of strains in the gut microbiome, and did not lead to significant differences in nGD values (Kruskal–Wallis test, $\chi^2 = 2.34$, $P = 0.31$; Extended Data Fig. 10a). nGD thresholds also did not significantly differ by phylum (Extended Data Fig. 10b), and those set in stool and oral samples were similar (median nGD difference = 0.006). If not limiting at 5% the fraction of unrelated individuals to share the same strain as a bound on a false discovery rate, the resulting percentile would only be of a median of 8.2% (range = [5.2–22.3%]) on these 38 SGBs (Supplementary Table 4). When using single metagenomic datasets instead of the five datasets we included to set the strain identity thresholds, often not enough longitudinal samples were available (<50 same-individual pairs) and some variation was observed (Extended Data Fig. 10c), which supports the use of the largest set of samples available.

Overall, the median SNV rate nGD thresholds corresponded to is 0.005, below the estimated >0.1% sequencing error rate by Illumina HiSeq and NovaSeq platforms[96] (Supplementary Table 4). The nGD thresholds correspond to a SNV rate of 0 for some SGBs ($n = 16$ out of 646—that is, 2.5%), mostly those encompassing very low genetic variation (for example, *B. animalis* SGB17278). In SGB trees containing MAGs of microorganisms obtained from fermented foods, we identified and discarded any strains with high similarity (≤0.0015 SNV rate as determined by PhyloPhlAn 3 (https://github.com/biobakery/phylophlan/wiki#mutation-rates-table)—that is, the number of positions that have nucleotide differences divided by the length of the alignment) to food MAGs (Supplementary Table 6). For *B. animalis* (SGB17278), 62 strains profiled in 7 public mouse metagenome datasets[73,75,97,98,99,100,101] were added to better assess its phylogenetic diversity. The trees produced by StrainPhlAn together with the SGB-specific nGD thresholds were used in StrainPhlAn4's strain_transmission.py script (-threshold argument)

(https://github.com/biobakery/MetaPhlAn/blob/master/metaphlan/utils/strain_transmission.py). Pairs of strains with pairwise nGD below the strain identity threshold were defined as strain-sharing events. Centred nGD is defined as the nGD divided by the median nGD in the phylogenetic tree. We opted for strain identity thresholds based on phylogenetic distances in contrast to SNV rates due to (1) the rather low coverage that we obtain for species in metagenomic samples even after passing our sequencing depth threshold (mean coverage = 7.2×, median = 0.69 and IQR = [0.14, 3.09]) that would add noise especially to SNV rate estimations; (2) the limited length of the marker gene alignment of some SGBs (mean trimmed alignment length = 74,348 nt, median = 70,879 and IQR = [42,513, 104,347]) that would make SNV rates rather unreliable; and (3) the valuable information on evolutionary models (for example, distinguishing synonymous from non-synonymous nucleotide changes) that is provided by phylogenetic trees.

We compared the new species-specific strain identity thresholds with the nGD = 0.1 threshold (that is, considering the lowest 10% phylogenetic distances to be between the same strains) used in some previous publications and StrainPhlAn versions prior to version 4 (refs. [9,32,102]). We found that while the previous threshold would produce a median 44% mother–infant strain-sharing rate—in contrast to the 50% strain-sharing rate we obtain here—the novel method yields a lower strain-sharing rate between infants and unrelated mothers, which are likely to be false positives: 3.5% versus 4%. This supports the better performance of the species-specific strain identity thresholds as they detect—at the same time—more strain-sharing events between matched mothers and infants and fewer strain-sharing events between unrelated mother–infant pairs.

To assess the reproducibility of the species-specific strain identity thresholds on additional unrelated data, we used independent datasets of patients undergoing faecal microbiome transplantation (FMT). As we used the publicly available metagenomic cohorts with no intervention and longitudinal sampling[20,22,27,28,31] to set the species-specific thresholds, we used for validation the completely independent FMT datasets as a distinct setting in which strain transmission can be expected. In FMT, part of the strains from a healthy donor are successfully transferred to a patient, while some strains from the donor's original sample remain after the intervention. We included 1,371 samples from 25 different cohorts of patients undergoing

FMT[103,104,105,106,107,108,109,110,111,112,113,114,115,116,117,118,119,120,121,122,123] that were analysed as part of a meta-analysis[124]. In this evaluation, similar to what we did in the set of longitudinal samples, we assessed the separation between the distribution of the nGD distances of strains from the same SGB in the two following situations: (1) the strains are from samples of the same individual or from a FMT donor and their recipient after the FMT, and (2) the strains are from samples belonging to different FMT triads (defined by the samples from the donor, those of the patient before FMT, and those of the patients after FMT). We performed this analysis for each of the 95 SGBs of our set that were also profiled in the Ianiro et al study. We considered as true positives pairwise phylogenetic distance (nGD) values between samples in (1) that were below the species-specific strain identity threshold (defined on the independent longitudinal datasets), false positives as those from (2) that were below the threshold, true negatives as those from (2) above the threshold, and false negatives as those from (1) above the threshold. We found that StrainPhlAn4 with the species-specific strain identity thresholds defined here performed very well in distinguishing strains in the same individual or FMT triad from different strains in different FMT triads: median recall = 0.97 and IQR = [0.95,0.99], precision = 0.72 [0.67,0.82], *F*-score = 0.97 [0.96,0.98] (Supplementary Table 35).

Assessment of person–person strain-sharing rates and SGB transmissibility

Person-to-person strain-sharing rates were calculated as the number of strains shared between two individuals divided by the number of shared SGBs profiled by StrainPhlAn (number of shared strains/number of shared SGBs). When multiple samples were available for an individual, detection of strain or SGB sharing at any time point was considered as the strain or SGB was shared. For a robust calculation, person-to-person strain-sharing rates were only assessed when at least ten SGBs were shared between two individuals. The same calculation was used to assess same-individual strain retention between two time points in longitudinal datasets. Strain acquisition rates by the offspring (Extended Data Fig. 6a) were defined as the proportion of strains profiled in the offspring that were shared with the mother, thus putatively originating from her. For a robust calculation, strain acquisition rates by the offspring were only assessed when at least ten SGBs were

shared between the mother and the offspring. As StrainPhlAn[36,38,39] profiles the dominant strain for each species, the total number of strains shared between two samples ranges between 0 and the total number of shared profiled SGBs, whereas strain-sharing rates and strain acquisition rates by the offspring are bound between 0 and 1.

SGB transmissibility was defined as the number of strain-sharing events detected for an SGB divided by the total potential number of strain-sharing events based on the presence of a strain-level profile by StrainPhlAn4. When multiple samples were available for an individual, detection of strain sharing at any time point was considered as the strain was shared. For a robust calculation, SGB transmissibility was only assessed on SGBs with at least ten potential strain-sharing events in multiple datasets, and with at least three potential strain-sharing events for single dataset calculations. To assess concordance of SGB transmissibility among datasets, Spearman's correlations (cor.test function in R (https://www.R-project.org/)) were performed between datasets with at least ten SGBs with assessed transmissibility. Highly transmitted SGBs were defined as those with SGB transmissibility >0.5 and significantly higher within-group than among-group transmissibility (Chi-squared tests, $P_{adj} < 0.05$). We found no significant association between SGB transmissibility and the length of the trimmed alignment (Spearman's test, $\rho = 0.06$, $P = 0.13$).

We assessed strain sharing across three main transmission modes: mother–infant (defined between mother and their offspring up to one year of age), household (defined as between cohabiting individuals), and intra-population (defined as that between non-cohabiting individuals in a population with no evidence of kinship).

Species-level beta diversity and ordination

For the appropriate analysis of microbiome compositional data, species-level abundance matrices obtained by MetaPhlAn were centred log ratio-transformed using the codaSeq.clr function in the CoDaSeq R package (v0.99.6)[125], using the minimum proportional abundance detected for each taxon for the imputation of zeros. A principal component analysis plot on Aitchison distance was produced with the ordinate and plot_ordination function in phyloseq (v1.28.0)[126], using one randomly selected sample per individual ($n = 4{,}840$ gut samples, $n = 2{,}069$ oral samples). To compare species-level similarity to strain-sharing rates, beta diversity metrics (Aitchison distance, Bray–Curtis dissimilarity, and Jaccard binary distance) computed with the vegan R package (v2.5–7) were converted to similarity indices (1 − (distance or dissimilarity)).

Strain–sharing networks

Unsupervised networks based on shared strains and species were visualized with R packages ggraph (v2.0.5), igraph (v1.2.6)[127], and tidygraph (v1.2.0) with stress layout, showing connections with ≥5 shared strains or ≥50 shared species (edges) among individuals (nodes).

Annotation of species phenotypic traits

Experimentally determined bacterial phenotypes were fetched from the Microbe Directory v2.0 (ref. [128]), and matched to kSGBs by NCBI taxonomic identifiers. Phenotypic traits that have previously been hypothesized to be linked with species transmissibility[3] were predicted for all SGBs using Traitar (version 1.1.12)[60] on the 50% core genes (genes present in 50% of genomes available in the expanded SGB database). Only annotations for which the phypat and the phypat + PGL classifiers (the second including additionally evolutionary information on phenotype gains and losses) annotations matched were kept. Associations between SGB transmissibility and microorganism phenotypes were assessed with

Wilcoxon rank-sum tests on the 25% most transmissible SGBs as compared to the 25% least transmissible ones.

Statistical analysis

Statistical analyses and graphical representations were performed in R using packages vegan (version 2.5–7), phyloseq (v1.28.0)[126], QuantPsyc (v1.5), ggplot2 (v3.3.3), ggpubr (v0.4.0) and corrplot (v0.84). Correction for multiple testing (Benjamini–Hochberg procedure, P_{adj}) was applied when appropriate and significance was defined at $P_{adj} < 0.05$. All tests were two-sided except where specified otherwise. The association between metadata variables and distance matrices was assessed by PERMANOVA with the adonis function in vegan. Differences between two groups were assessed with Wilcoxon rank-sum tests. For more than two groups, the Kruskal–Wallis test with post hoc Dunn tests was used. Correlations were assessed with Spearman's tests. To assess correlations between variables while partialling out potential confounders, GLMs were fitted with the glm R function (Gaussian, link = identity). Standardized GLM regression coefficients were calculated using the lm.beta R function (QuantPsyc R package). The significance was assessed by performing log likelihood (Chi-squared) tests on nested GLMs.

Ethical compliance

All study procedures are compliant with all relevant ethical regulations. The procedures were performed in compliance with the Declaration of Helsinki. Ethical approval of the Argentina cohort was granted by the Ethics and Safety committee (CEySTE), CCT Santa Fe, Argentina (29112019). The Colombia cohort was approved by the Research Bioethics committee, Universidad Metropolitana, Colombia (NIT 890105361-5). The China_1 dataset research protocol was approved by the Ethics Committee of Shanghai Tenth Hospital, Tongji University School of Medicine (SHSY-IEC-pap-18-1), and China_2 was approved by the Ethics committee of the Health Science Center, Xi'an Jiaotong University, China (2016-114). The Guinea-Bissau study was approved by the Health Ethics National Committee (Comitê Nacional da Ética na Saude), Ministry of Public

Health, Guinea-Bissau (076/CNES/INASA/2017) and by the London School of Hygiene and Tropical Medicine Ethics Committee (reference number 22898). The Italy_1 dataset research protocol was approved by the Ethics Committee of Santa Chiara Hospital, Trento, Italy (51082283, 30 July 2014) and the Ethics Committee of the University of Trento, Italy, and Italy_2 by the Liguria Regional Ethics Committee, Italy (006/2019). Ethical approval for the USA dataset was granted by Western IRB (https://www.wcgirb.com/), with protocol tracking number WIRB20151664. Written informed consent was obtained from all adult participants and from parents of non-adult participants.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Shotgun metagenomics sequencing data of the Argentina, Colombia, China_2, Guinea-Bissau, Italy_1 and USA datasets are available at the European Nucleotide Archive under accession number PRJEB45799. The sequencing data of the China_1 dataset is available on the NCBI Sequence Read Archive database with accession PRJNA613947. The sequencing data of the Italy_2 dataset is on the NCBI Sequence Read Archive database with accession PRJNA716780. Metadata are available in Supplementary Table 2 and in the latest release of curatedMetagenomicData[61].

Code availability

All the software and thresholds developed and used in this study are available in the MetaPhlAn4 package[39] (which includes StrainPhlAn4 and the script for strain transmission inference with the species-specific strain identity thresholds), available at http://segatalab.cibio.unitn.it/tools/metaphlan with the open-source code at https://github.com/biobakery/MetaPhlAn. It is also available via Bioconda (https://anaconda.org/bioconda/metaphlan) and PIP

(https://pypi.org/project/MetaPhlAn). A tutorial describing the procedure we followed to assess strain sharing is available at https://github.com/biobakery/MetaPhlAn/wiki/Strain-Sharing-Inference.

References

1. Falony, G. et al. Population-level analysis of gut microbiome variation. *Science* **352**, 560–564 (2016).

2. Asnicar, F. et al. Microbiome connections with host metabolism and habitual diet from 1,098 deeply phenotyped individuals. *Nat. Med.* **27**, 321–332 (2021).

3. Browne, H. P., Neville, B. A., Forster, S. C. & Lawley, T. D. Transmission of the gut microbiota: spreading of health. *Nat. Rev. Microbiol.* **15**, 531–543 (2017).

4. Robinson, C. D., Bohannan, B. J. & Britton, R. A. Scales of persistence: transmission and the microbiome. *Curr. Opin. Microbiol.* **50**, 42–49 (2019).

5. Finlay, B. B. & CIFAR Humans and the Microbiome. Are noncommunicable diseases communicable? *Science* **367**, 250–251 (2020).

6. Stewart, C. J. et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* **562**, 583–588 (2018).

7. Chen, L. et al. The long-term genetic stability and individual specificity of the human gut microbiome. *Cell* **184**, 2302–2315 (2021).

8. David, L. A. et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014).

9. Ferretti, P. et al. Mother-to-infant microbial transmission from different body sites shapes the developing infant gut microbiome. *Cell Host Microbe* **24**, 133–145.e5 (2018).

10. Asnicar, F. et al. Studying vertical microbiome transmission from mothers to infants by strain-level metagenomic profiling. *mSystems* **2**, e00164–16 (2017).

11. Korpela, K. et al. Selective maternal seeding and environment shape the human gut microbiome. *Genome Res.* **28**, 561–568 (2018).

12. Yassour, M. et al. Strain-level analysis of mother-to-child bacterial transmission during the first few months of life. *Cell Host Microbe* **24**, 146–154.e4 (2018).

13. Nayfach, S., Rodriguez-Mueller, B., Garud, N. & Pollard, K. S. An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. *Genome Res.* **26**, 1612–1625 (2016).

14. Podlesny, D. & Fricke, W. F. Strain inheritance and neonatal gut microbiota development: a meta-analysis. *Int. J. Med. Microbiol.* **311**, 151483 (2021).

15. Moeller, A. H. et al. Social behavior shapes the chimpanzee pan-microbiome. *Sci. Adv.* **2**, e1500997 (2016).

16. Sarkar, A. et al. Microbial transmission in animal social networks and the social microbiome. *Nat. Ecol. Evol.* **4**, 1020–1035 (2020).

17. Brito, I. L. et al. Transmission of human-associated microbiota along family and social networks. *Nat. Microbiol.* **4**, 964–971 (2019).

18. Segata, N. On the road to strain-resolved comparative metagenomics. *mSystems* **3**, e00190–17 (2018).

19. Van Rossum, T., Ferretti, P., Maistrenko, O. M. & Bork, P. Diversity within species: interpreting strains in microbiomes. *Nat. Rev. Microbiol.* **18**, 491–506 (2020).

20. Nielsen, H. B. et al. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using

reference genomes. *Nat. Biotechnol.* **32**, 822–828 (2014).

21. Bäckhed, F. et al. Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe* **17**, 690–703 (2015).

22. Louis, S., Tappu, R.-M., Damms-Machado, A., Huson, D. H. & Bischoff, S. C. Characterization of the gut microbial community of obese patients following a weight-loss intervention using whole metagenome shotgun sequencing. *PLoS ONE* **11**, e0149564 (2016).

23. Pehrsson, E. C. et al. Interconnected microbiomes and resistomes in low-income human habitats. *Nature* **533**, 212–216 (2016).

24. Brito, I. L. et al. Mobile genes in the human microbiome are structured from global to individual scales. *Nature* **535**, 435–439 (2016).

25. Xie, H. et al. Shotgun metagenomics of 250 adult twins reveals genetic and environmental impacts on the gut microbiome. *Cell Syst.* **3**, 572–584.e3 (2016).

26. Chu, D. M. et al. Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. *Nat. Med.* **23**, 314–326 (2017).

27. Costea, P. I. et al. Subspecies in the global human gut microbiome. *Mol. Syst. Biol.* **13**, 960 (2017).

28. Mehta, R. S. et al. Stability of the human faecal microbiome in a cohort of adult men. *Nat. Microbiol.* **3**, 347–355 (2018).

29. Wampach, L. et al. Birth mode is associated with earliest strain-conferred gut microbiome functions and immunostimulatory potential. *Nat. Commun.* **9**, 5091 (2018).

30. Pasolli, E. et al. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell* **176**, 649–662.e20 (2019).

31. Lloyd-Price, J. et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **569**, 655–662 (2019).

32. Shao, Y. et al. Stunted microbiota and opportunistic pathogen colonization in caesarean-section birth. *Nature* **574**, 117–121 (2019).

33. Visconti, A. et al. Interplay between the human gut microbiome and host metabolism. *Nat. Commun.* **10**, 4505 (2019).

34. Tett, A. et al. The *Prevotella copri* complex comprises four distinct clades underrepresented in westernized populations. *Cell Host Microbe* **26**, 666–679.e7 (2019).

35. Lloyd-Price, J. et al. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature* **550**, 61–66 (2017).

36. Truong, D. T., Tett, A., Pasolli, E., Huttenhower, C. & Segata, N. Microbial strain-level population structure and genetic diversity from metagenomes. *Genome Res.* **27**, 626–638 (2017).

37. Albanese, D. & Donati, C. Strain profiling and epidemiology of bacterial species from metagenomic sequencing. *Nat. Commun.* **8**, 2260 (2017).

38. Beghini, F. et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife* **10**, e65088 (2021).

39. Blanco-Miguez, A. et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species with MetaPhlAn 4. Preprint at *bioRxiv* https://doi.org/10.1101/2022.08.22.504593 (2022).

40. Pasolli, E. et al. Large-scale genome-wide analysis links lactic acid bacteria from food with the gut microbiome. *Nat. Commun.* **11**, 2610 (2020).

41. Website. Lou et al. Infant gut strain persistence is associated with maternal origin, phylogeny, and functional potential including surface

adhesion and iron acquisition. *Cell Host Microbe* https://doi.org/10.2139/ssrn.3778932 (2021).

42. Jenni, O. G., Chaouch, A., Caflisch, J. & Rousson, V. Infant motor milestones: poor predictive value for outcome of healthy children. *Acta Paediatr.* **102**, e181–e184 (2013).

43. Segata, N. Gut microbiome: westernization and the disappearance of intestinal diversity. *Curr. Biol.* **25**, R611–R613 (2015).

44. Sonnenburg, E. D. et al. Diet-induced extinctions in the gut microbiota compound over generations. *Nature* **529**, 212–215 (2016).

45. Yatsunenko, T. et al. Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227 (2012).

46. Rutayisire, E., Huang, K., Liu, Y. & Tao, F. The mode of delivery affects the diversity and colonization pattern of the gut microbiota during the first year of infants' life: a systematic review. *BMC Gastroenterol.* **16**, 86 (2016).

47. Song, S. J. et al. Cohabiting family members share microbiota with one another and with their dogs. *eLife* **2**, e00458 (2013).

48. Qian, Y. et al. Gut metagenomics-derived genes as potential biomarkers of Parkinson's disease. *Brain* **143**, 2474–2489 (2020).

49. Rothschild, D. et al. Environment dominates over host genetics in shaping human gut microbiota. *Nature* **555**, 210–215 (2018).

50. Obregon-Tito, A. J. et al. Subsistence strategies in traditional societies distinguish gut microbiomes. *Nat. Commun.* **6**, 6505 (2015).

51. Vangay, P. et al. US immigration westernizes the human gut microbiome. *Cell* **175**, 962–972.e10 (2018).

52. Bergey, D. H. et al. *Bergey's Manual of Systematic Bacteriology* (Lippincott Raven, 1989).

53. Uriot, O. et al. *Streptococcus thermophilus*: From yogurt starter to a new promising probiotic candidate? *J. Funct. Foods* **37**, 74–89 (2017).

54. Kort, R. et al. Shaping the oral microbiota through intimate kissing. *Microbiome* **2**, 41 (2014).

55. Jo, R. et al. Comparison of oral microbiome profiles in 18-month-old infants and their parents. *Sci. Rep.* **11**, 861 (2021).

56. Hildebrand, F. et al. Dispersal strategies shape persistence and evolution of human gut bacteria. *Cell Host Microbe* **29**, 1167–1176.e9 (2021).

57. Mahnert, A. et al. Man-made microbial resistances in built environments. *Nat. Commun.* **10**, 968 (2019).

58. Guittar, J., Shade, A. & Litchman, E. Trait-based community assembly and succession of the infant gut microbiome. *Nat. Commun.* **10**, 512 (2019).

59. Josenhans, C. & Suerbaum, S. The role of motility as a virulence factor in bacteria. *Int. J. Med. Microbiol.* **291**, 605–614 (2002).

60. Weimann, A. et al. From genomes to phenotypes: Traitar, the microbial trait analyzer. *mSystems* **1**, e00101–16 (2016).

61. Pasolli, E. et al. Accessible, curated metagenomic data through ExperimentHub. *Nat. Methods* **14**, 1023–1024 (2017).

62. Benezra, A. Race in the microbiome. *Sci. Technol. Hum. Values* **45**, 877–902 (2020).

63. Xu, Q. et al. Metagenomic and metabolomic remodeling in nonagenarians and centenarians and its association with genetic and socioeconomic factors. *Nat. Aging* **2**, 438–452 (2022).

64. Farrant, O. et al. Prevalence, risk factors and health consequences of soil-transmitted helminth infection on the Bijagos Islands, Guinea

Bissau: a community-wide cross-sectional study. *PLoS Negl. Trop. Dis.* **14**, e0008938 (2020).

65. The Human Microbiome Project Consortium. A framework for human microbiome research. *Nature* **486**, 215–221 (2012).

66. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

67. Wu, L. et al. A cross-sectional study of compositional and functional profiles of gut microbiota in Sardinian centenarians. *mSystems* **4**, e00325–19 (2019).

68. Manara, S. et al. Microbial genomes from non-human primate gut metagenomes expand the primate-associated bacterial tree of life with over 1000 novel species. *Genome Biol.* **20**, 299 (2019).

69. Stewart, R. D. et al. Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nat. Biotechnol.* **37**, 953–961 (2019).

70. Xiao, L. et al. A catalog of the mouse gut metagenome. *Nat. Biotechnol.* **33**, 1103–1108 (2015).

71. Sharpton, T. et al. Development of inflammatory bowel disease is linked to a longitudinal restructuring of the gut metagenome in mice. *mSystems* **2**, e00036–17 (2017).

72. Xiao, L. et al. High-fat feeding rather than obesity drives taxonomical and functional changes in the gut microbiota in mice. *Microbiome* **5**, 43 (2017).

73. Hebbandi Nanjundappa, R. et al. A gut microbial mimic that hijacks diabetogenic autoreactivity to suppress colitis. *Cell* **171**, 655–667.e17 (2017).

74. Rosshart, S. P. et al. Wild mouse gut microbiota promotes host fitness and improves disease resistance. *Cell* **171**, 1015–1028.e13 (2017).

75. Rosshart, S. P. et al. Laboratory mice born to wild mice have natural microbiota and model human immune responses. *Science* **365**, eaaw4361 (2019).

76. Kreznar, J. H. et al. Host genotype and gut microbiome modulate insulin secretion and diet-induced metabolic phenotypes. *Cell Rep.* **18**, 1739–1750 (2017).

77. Lagkouvardos, I. et al. The mouse intestinal bacterial collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota. *Nat. Microbiol.* **1**, 16131 (2016).

78. Riva, A. et al. A fiber-deprived diet disturbs the fine-scale spatial architecture of the murine colon microbiome. *Nat. Commun.* **10**, 4366 (2019).

79. Reyes, A., Wu, M., McNulty, N. P., Rohwer, F. L. & Gordon, J. I. Gnotobiotic mouse model of phage–bacterial host dynamics in the human gut. *Proc. Natl Acad. Sci. USA* **110**, 20236–20241 (2013).

80. Lesker, T. R. et al. An integrated metagenome catalog reveals new insights into the murine gut microbiome. *Cell Rep.* **30**, 2909–2922.e6 (2020).

81. Blacher, E. et al. Potential roles of gut microbiome and metabolites in modulating ALS in mice. *Nature* **572**, 474–480 (2019).

82. Ni, Y. et al. A metagenomic study of the preventive effect of *Lactobacillus rhamnosus* GG on intestinal polyp formation in $Apc^{Min/+}$ mice. *J. Appl. Microbiol.* **122**, 770–784 (2017).

83. Hughes, E. R. et al. Microbial respiration and formate oxidation as metabolic signatures of inflammation-associated dysbiosis. *Cell Host Microbe* **21**, 208–219 (2017).

84. Parks, D. H. et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* **2**,

1533–1542 (2017).

85. Li, D. et al. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* **102**, 3–11 (2016).

86. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).

87. Kang, D. D. et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).

88. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).

89. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

90. Suzek, B. E. et al. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **31**, 926–932 (2015).

91. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).

92. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).

93. Mirdita, M. et al. Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Res.* **45**, D170–D176 (2017).

94. Asnicar, F. et al. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nat. Commun.* **11**, 2500 (2020).

95. Ondov, B. D. et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **17**, 132 (2016).

96. Stoler, N. & Nekrutenko, A. Sequencing error profiles of Illumina sequencing instruments. *Genomics Bioinform.* **3**, lqab019 (2021).

97. Kim, M.-S. & Bae, J.-W. Lysogeny is prevalent and widely distributed in the murine gut microbiota. *ISME J.* **12**, 1127–1141 (2018).

98. Kibe, R. et al. Upregulation of colonic luminal polyamines produced by intestinal microbiota delays senescence in mice. *Sci Rep.* **4**, 4548 (2014).

99. Jovel, J. et al. Characterization of the gut microbiome using 16s or shotgun metagenomics. *Front. Microbiol.* **7**, 459 (2016).

100. Fabbiano, S. et al. Functional gut microbiota remodeling contributes to the caloric restriction-induced metabolic improvements. *Cell Metab.* **28**, 907–921.e7 (2018).

101. Yang, H. et al. Truncation of mutant huntingtin in knock-in mice demonstrates exon1 huntingtin is a key pathogenic form. *Nat. Commun.* **11**, 2582 (2020).

102. Valles-Colomer, M. et al. Variation and transmission of the human gut microbiota across multiple familial generations. *Nat. Microbiol.* **7**, 87–96 (2022).

103. Aggarwala, V. et al. Precise quantification of bacterial strains after fecal microbiota transplantation delineates long-term engraftment and explains outcomes. *Nat. Microbiol.* **6**, 1309–1318 (2021).

104. Baruch, E. N. et al. Fecal microbiota transplant promotes response in immunotherapy-refractory melanoma patients. *Science* **371**, 602–609

(2021).

105. Bar-Yoseph, H. et al. Oral capsulized fecal microbiota transplantation for eradication of carbapenemase-producing *Enterobacteriaceae* colonization with a metagenomic perspective. *Clin. Infect. Dis.* **73**, e166–e175 (2021).

106. Damman, C. J. et al. Low level engraftment and improvement following a single colonoscopic administration of fecal microbiota to patients with ulcerative colitis. *PLoS ONE* **10**, e0133925 (2015).

107. Davar, D. et al. Fecal microbiota transplant overcomes resistance to anti-PD-1 therapy in melanoma patients. *Science* **371**, 595–602 (2021).

108. Goll, R. et al. Effects of fecal microbiota transplantation in subjects with irritable bowel syndrome are mirrored by changes in gut microbiome. *Gut Microbes* **12**, 1794263 (2020).

109. Hourigan, S. K. et al. Fecal transplant in children with *Clostridioides difficile* gives sustained reduction in antimicrobial resistance and potential pathogen burden. *Open Forum Infect. Dis.* **6**, ofz379 (2019).

110. Ianiro, G. et al. Faecal microbiota transplantation for the treatment of diarrhoea induced by tyrosine-kinase inhibitors in patients with metastatic renal cell carcinoma. *Nat. Commun.* **11**, 4333 (2020).

111. Kong, L. et al. Linking strain engraftment in fecal microbiota transplantation with maintenance of remission in Crohn's disease. *Gastroenterology* **159**, 2193–2202 (2020).

112. Koopen, A. M. et al. Effect of fecal microbiota transplantation combined with Mediterranean diet on insulin sensitivity in subjects with metabolic syndrome. *Front. Microbiol.* **12**, 662159 (2021).

113. Kumar, R. et al. Identification of donor microbe species that colonize and persist long term in the recipient after fecal transplant for recurrent *Clostridium difficile*. *NPJ Biofilms Microbiomes* **3**, 12 (2017).

114. Leo, S. et al. Metagenomic characterization of gut microbiota of carriers of extended-spectrum beta-lactamase or carbapenemase-producing *Enterobacteriaceae* following treatment with oral antibiotics and fecal microbiota transplantation: results from a multicenter randomized trial. *Microorganisms* **8**, 941 (2020).

115. Li, S. S. et al. Durable coexistence of donor and recipient strains after fecal microbiota transplantation. *Science* **352**, 586–589 (2016).

116. Moss, E. L. et al. Long-term taxonomic and functional divergence from donor bacterial strains following fecal microbiota transplantation in immunocompromised patients. *PLoS ONE* **12**, e0182585 (2017).

117. Podlesny, D. & Fricke, W. F. Microbial strain engraftment, persistence and replacement after fecal microbiota transplantation. Preprint at *bioRxiv* https://doi.org/10.1101/2020.09.29.20203638 (2020).

118. Smillie, C. S. et al. Strain tracking reveals the determinants of bacterial engraftment in the human gut following fecal microbiota transplantation. *Cell Host Microbe* **23**, 229–240.e5 (2018).

119. Suskind, D. L. et al. Fecal microbial transplant effect on clinical outcomes and fecal microbiome in active Crohn's disease. *Inflamm. Bowel Dis.* **21**, 556–563 (2015).

120. Vaughn, B. P. et al. Increased intestinal microbial diversity following fecal microbiota transplant for active Crohn's disease. *Inflamm. Bowel Dis.* **22**, 2182–2190 (2016).

121. Verma, S. et al. Identification and engraftment of new bacterial strains by shotgun metagenomic sequence analysis in patients with recurrent *Clostridioides difficile* infection before and after fecal microbiota transplantation and in healthy human subjects. *PLoS ONE* **16**, e0251590 (2021).

122. Watson, A. R. et al. Adaptive ecological processes and metabolic independence drive microbial colonization and resilience in the human

gut. Preprint at *bioRxiv* https://doi.org/10.1101/2021.03.02.433653 (2021).

123. Zhao, H.-J. et al. The efficacy of fecal microbiota transplantation for children with Tourette syndrome: a preliminary study. *Front. Psychiatry* **11**, 554441 (2020).

124. Ianiro, G. et al. Variability of strain engraftment and predictability of microbiome composition after fecal microbiota transplantation across different diseases. *Nat. Med.* **28**, 1913–1923 (2022).

125. Gloor, G. B. & Reid, G. Compositional analysis: a valid approach to analyze microbiome high-throughput sequencing data. *Can. J. Microbiol.* **62**, 692–703 (2016).

126. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217 (2013).

127. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal Complex Syst.* **1695**1–9 (2006).

128. Sierra, M. A. et al. The Microbe Directory v2.0: an expanded database of ecological and phenotypical features of microbes. Preprint at *bioRxiv* https://doi.org/10.1101/2019.12.20.860569 (2019).

Acknowledgements

We thank all study participants for their commitment; N. Volfovsky, P. Feliciano and A. Packer from the Simons Foundations for their kind support with the SPARK collection data; the LaBSSAH—CIBIO Next Generation Sequencing Facility of the University of Trento (V. de Sanctis, R. Bertorelli, P. Cavallerio and C. Valentini) for sequencing the metagenomic libraries. This work was supported by the European Research Council (ERC-STG project MetaPG-716575 and ERC-CoG microTOUCH-101045015) to N.S. and by EMBO ALTF 593–2020 to M.V.-C. The work was also partially supported by MIUR 'Futuro in Ricerca' (grant no. RBFR13EWWI_001) to

N.S., by the European H2020 programme (ONCOBIOME-825410 project, MASTER-818368 project, and IHMCSA-964590) to N.S., by the National Cancer Institute of the National Institutes of Health (1U01CA230551) to N.S., by the Premio Internazionale Lombardia e Ricerca 2019 to N.S., by the Simons Foundation (award ID 648614) to E.D. and N.S., and by the European Research Council (ERC-STG project Mami-639226) to M.C.C.

Author information

Authors and Affiliations

1. Department CIBIO, University of Trento, Trento, Italy

   Mireia Valles-Colomer, Aitor Blanco-Míguez, Paolo Manghi, Francesco Asnicar, Leonard Dubois, Davide Golzato, Federica Armanini, Fabio Cumbo, Kun D. Huang, Serena Manara, Giulia Masetti, Federica Pinto, Michal Punčochář, Liviana Ricci, Moreno Zolfo, Marta Selma-Royo, Michele Filosi, Adrian Tett, Enrico Domenici & Nicola Segata

2. Department of Experimental Oncology, IEO European Institute of Oncology IRCCS, Milan, Italy

   Elisa Piperni & Nicola Segata

3. Clinical Research Department, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK

   Olivia Farrant, Adriana Goncalves & Anna Last

4. Institute of Agrochemistry and Food Technology-National Research Council (IATA-CSIC), Paterna, Valencia, Spain

   Marta Selma-Royo & Maria Carmen Collado

5. Instituto de Lactología Industrial (CONICET-UNL), Facultad de Ingeniería Química, Universidad Nacional del Litoral, Santa Fe, Argentina

Ana G. Binetti

6. Grupo de Investigación Alimentación y Comportamiento Humano, Universidad Metropolitana, Barranquilla, Colombia

Jimmy E. Becerra

7. School of Public Health, Health Science Center, Xi’an Jiaotong University, Xi’an, China

Bei Han

8. National Institute for Medical Research, Tanga Medical Research Centre, Tanga, Tanzania

John Lusingu

9. Kumasi Centre for Collaborative Research in Tropical Medicine, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

John Amuasi

10. Oncology Unit, IRCCS Istituto Giannina Gaslini, Genoa, Italy

Loredana Amoroso

11. Department of Twin Research and Genetic Epidemiology, King’s College London, London, UK

Alessia Visconti, Claire M. Steves, Mario Falchi & Tim D. Spector

12. Centre for Microbiology and Environmental Systems Science, University of Vienna, Vienna, Austria

Adrian Tett

13. Shanghai Tenth People’s Hospital, Tongji University School of Medicine, Shanghai, China

Qian Xu, Nan Qin & Huanlong Qin

14. Realbio Genomics Institute, Shanghai, China

Qian Xu & Nan Qin

15. Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany

Jürgen May & Daniel Eibach

16. Laboratory of Experimental Therapies in Oncology, IRCCS Istituto Giannina Gaslini, Genoa, Italy

Maria Valeria Corrias & Mirco Ponzoni

17. Department of Agricultural Sciences, University of Naples ‘Federico II’, Portici, Italy

Edoardo Pasolli

18. Centre for Computational and Systems Biology (COSBI), Microsoft Research Foundation, Rovereto, Italy

Enrico Domenici

Contributions

M.V.-C. and N.S. conceived and designed the study. A.B.-M., F. Asnicar and F.C. contributed software developments on StrainPhlAn, PhyloPhlAn and the expanded SGB database. P.M., D.G., F. Armanini, F.C., K.D.H., S.M., G.M., F.P., E. Piperni, M. Punčochář, L.R., M.Z., O.F., A.G., M.S.-R., A.G.B., J.E.B., B.H., J.L., J.A., L.A., A.V., C.M.S., M. Falchi, M. Filosi, A.T., A.L., Q.X., N.Q., H.Q., J.M., D.E., M.V.-C., M. Ponzoni, E. Pasolli, T.D.S., E.D. and M.C.C. participated in cohort recruitment and/or

contributed to metadata and metagenomic data acquisition. L.D. and M.V.-C. developed the tutorial for strain-sharing inference. F. Armanini performed DNA extraction and shotgun sequencing library preparation on the novel datasets. F. Asnicar, P.M. and M.Z. pre-processed and quality-controlled the metagenomic datasets. M.V.-C. and N.S. performed the analysis, interpreted the data and wrote the manuscript. All authors provided critical revision of the manuscript and approved the final version for submission.

Corresponding authors

Correspondence to Mireia Valles-Colomer or Nicola Segata.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Data overview.

A) Species-level ordination (PCoA on Aitchison distance, N = 2,069 samples) reflecting the overall microbiome diversity spanned by the oral microbiome samples considered. Samples are coloured by country, while shapes depict age. **B**) Colour code of the samples in the phylogenetic tree in Fig. 1c, representing the datasets they belong to.

Extended Data Fig. 2 Strain sharing workflow.

Workflow used to assess strain sharing in the current manuscript.

Extended Data Fig. 3 Species-specific operational definitions of strain.

Comparison of same-individual (green) to unrelated individual (purple) genetic distance comparisons for the 25 most prevalent SGBs in gut metagenome longitudinal datasets. Strain identity thresholds were set as the Youden's index (black dashed line) or as the 5th percentile of the unrelated individual comparisons (red dashed line) when the first was above 5% (e.g. *Parabacteroides merdae* [SGB1949]). Centred nGD: normalised phylogenetic distance divided by the median nGD of the phylogenetic tree. The N in each histogram corresponds to the number of same-individual comparisons in which each SGB was profiled at strain-level.

Extended Data Fig. 4 Phylogenetic trees of species containing strains found in food.

A) Phylogeny of *Bifidobacterium animalis* (SGB17278) produced with StrainPhlAn (*Methods*) including strains reconstructed from human gut metagenomes, from mice samples (grey dots) and MAGs reconstructed from fermented food 32 (yellow dots). Differently from strains found in mice, 94% of human-derived strains are at ≤0.0015 single nucleotide variation (SNV) rate to MAGs obtained from fermented food (*Methods*), suggesting that the presence of this species in humans is associated with consumption of commercial dietary products, and were consequently excluded from further analyses (horizontal grey bars). **B**) Phylogeny of *Streptococcus thermophilus-salivarius-vestibularis* (SGB8002) produced

with StrainPhlAn (*Methods*) including strains reconstructed from human gut metagenomes together with MAGs reconstructed from fermented food 32 (yellow dots), suggesting only a subset of strains found in the human gut is associated with fermented food intake. Only the leaves in the enlarged subtree ("Fermented food subtree") were at ≤0.0015 single nucleotide variation (SNV) rate to MAGs obtained from fermented food (Methods) and were consequently excluded from further analyses.

Extended Data Fig. 5 Strain and species-level similarity across relationships.

A) Gut microbiome strain sharing rates and species-level similarity metrics (Aitchison similarity, Bray-Curtis similarity, and Jaccard binary similarity) between individuals in the same household ("within household") as compared to unrelated non-cohabiting individuals in different villages of the same population ("within population") and individuals in different populations ("interpopulation"). For comparability with strain sharing rates, species-level comparisons are depicted as similarity indices (1 - distance or dissimilarity). All comparisons are significant (Padj<0.05, Kruskal-Wallis tests with Post-hoc Dunn tests, Table S8). The social-distance based gradient followed by strain sharing rates is notably stronger than that observed by species-level similarity metrics (Table S8). Boxes: lower and upper quartiles, middle line: median; whiskers: 1.5 × IQR. **B**) Oral microbiome strain sharing rates and species-level similarity metrics (Aitchison, Bray-Curtis, and Jaccard binary similarities) between individuals in the same household ("within household") as compared to unrelated non-cohabiting individuals in different villages of the same population ("within population") and individuals in different populations ("interpopulation"). For comparability with strain sharing rates, species-level comparisons are depicted as similarity indices (1 - distance or dissimilarity). All comparisons are significant (Padj<0.05, Kruskal-Wallis tests with Post-hoc Dunn tests, Table S28). Boxes: lower and upper quartiles, middle line: median; whiskers: 1.5 × IQR.

Extended Data Fig. 6 Mother to offspring gut microbiome transmission.

A) Strain acquisition rates by the offspring tend to decrease as a function of the offspring's age. Strain acquisition rates by the offspring are defined as the proportion of strains profiled in the offspring that are shared with their mother, computed in 17 datasets from 14 different countries across pre-defined age categories. Kruskal-Wallis test, Chi2=65, P = 3.57e-12, Post-hoc Dunn tests, NS corresponds to Padj≥0.05, all other comparisons are significant (Table S10). Boxes: lower and upper quartiles, middle line: median; whiskers: 1.5 × IQR. Novel datasets are highlighted with asterisks. **B**) Strain sharing rates between senior individuals and their non-cohabiting mothers as compared to strain sharing rates between unrelated mother-offspring pairs. Wilcoxon rank-sum test, N = 17,177, r = 0.09, P = 4.1e-35. Boxes: lower and upper quartiles, middle line: median; whiskers: 1.5 × IQR. **C**) Observed richness (number of SGBs detected with MetaPhlAn) in age categories of offspring from Westernized as compared to non-Westernized populations. Wilcoxon rank-sum tests, N = 721, ***Padj <0.001 and **Padj<0.01, Table S11. Boxes: lower and upper quartiles, middle line: median; whiskers: 1.5 × IQR. **D**) Mother-offspring strain sharing rates in age categories of offspring delivered by C-section as compared to vaginally-delivered offspring. Wilcoxon rank-sum tests, **Padj<0.01, NS Padj≥0.05, Table S14. Boxes: lower and upper quartiles, middle line: median; whiskers: 1.5 × IQR.

Extended Data Fig. 7 Gut microbiome strain sharing among adult twins.

Dizygotic and monozygotic twin gut microbiome strain sharing rates after decades since cohabitation. Wilcoxon rank-sum tests, N = 708, **Padj<0.01, *Padj<0.05, NS Padj≥0.05, Table S19. Boxes: lower and upper quartiles, middle line: median; whiskers: 1.5 × IQR.

Extended Data Fig. 8 Gut microbiome species and strain sharing among individuals.

A) Density distributions of gut microbiome strain sharing rates between household members (within household), individuals in different households in the same village (within village), individuals in different villages of the

same population (within population), and in different populations (interpopulation). **B**) Gut microbiome species sharing unsupervised network of household datasets. Line width is proportional to the number of shared species. Only connections with ≥50 shared species are shown.

Extended Data Fig. 9 Highly-transmitted SGBs in oral samples.

Same-family (green) to different-family (purple) genetic distance comparisons for the three SGBs consistently and significantly highly-transmitted in oral metagenomes. Strain identity thresholds were set as the 3rd percentile of the unrelated individual comparisons (dashed line).

Extended Data Fig. 10 Assessment of strain identity thresholds.

A) Centred nGD (normalised phylogenetic distance divided by the median nGD of the phylogenetic tree) used as a threshold for strain identity (corresponding to the percentiles of interindividual distributions) by strain definition used, for the 646 SGBs profiled in stool samples. The different percentiles do not result in significant differences in nGD values (Kruskal-Wallis test, Chi2=2.34, P = 0.31). Boxes: lower and upper quartiles, middle line: median; whiskers: 1.5 × IQR. **B**) Distribution of centred nGD thresholds (normalised phylogenetic distance divided by the median nGD of the phylogenetic tree) by phylum, showing lack of statistically-significant association (Kruskal-Wallis test, Chi2=6.6, P = 0.25). Boxes: lower and upper quartiles, middle line: median; whiskers: 1.5 × IQR. **C**) Strain identity thresholds (percentile of interindividual nGD distribution) calculated for each of the SGBs prevalent in longitudinal datasets (N = 145 SGBs profiled in at least 50 same-individual pairs) calculated on single datasets compared to the threshold used in the study (determined on all samples).

Supplementary information

Supplementary Information

This file contains a guide to Supplementary Tables 1–35 (tables supplied separately) and a link to a tutorial describing the procedure followed to assess strain sharing.

Reporting Summary

Supplementary Tables

Supplementary Tables 1–35: see Supplementary Information document for full descriptions.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Valles-Colomer, M., Blanco-Míguez, A., Manghi, P. *et al.* The person-to-person transmission landscape of the gut and oral microbiomes. *Nature* **614**, 125–135 (2023). https://doi.org/10.1038/s41586-022-05620-1

- Received: 14 July 2021
- Accepted: 02 December 2022
- Published: 18 January 2023
- Issue Date: 02 February 2023
- DOI: https://doi.org/10.1038/s41586-022-05620-1

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article is cited by

- **How our microbiome is shaped by family, friends and even neighbours**
    - Ewen Callaway

  Nature (2023)

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05620-1



Dendritic cells rise and shine

21 December 2022

Joseph Willson



Breast and pancreatic cancer interrupt IRF8-dependent dendritic cell development to overcome immune surveillance

28 March 2018

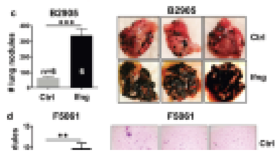
Melissa A. Meyer, John M. Baer, … David G. DeNardo



The reciprocal interplay between TNFα and the circadian clock impacts on cell proliferation and migration in Hodgkin lymphoma cells

31 July 2018

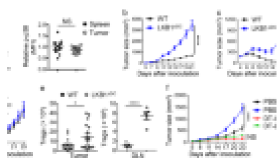
Mónica Abreu, Alireza Basti, … Angela Relógio



Interferon-gamma signaling promotes melanoma progression and metastasis

03 December 2022

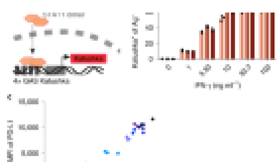
Bo Zhou, Jayati Basu, … M. Raza Zaidi



LKB1 orchestrates dendritic cell metabolic quiescence and anti-tumor immunity

25 March 2019

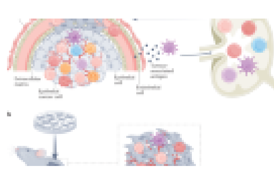
Yanyan Wang, Xingrong Du, … Hongbo Chi



Long-distance modulation of bystander tumor cells by CD8+ T-cell-secreted IFN-γ

09 March 2020

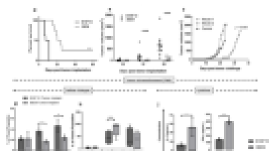
Mirjam E. Hoekstra, Laura Bornes, … Ton N. M. Schumacher



[TGFβ control of immune responses in cancer: a holistic immuno-oncology perspective]

15 November 2022

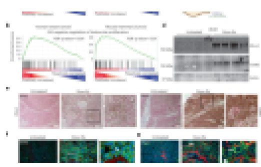
Briana G. Nixon, Shengyu Gao, … Ming O. Li



[Inducing Tumor Suppressive Microenvironments through Genome Edited CD47−/− Syngeneic Cell Vaccination]

27 December 2019

Subhadra Jayaraman Rukmini, Huanjing Bi, … Kaiming Ye



[Breast cancer cells survive chemotherapy by activating targetable immune-modulatory programs characterized by PD-L1 or CD80]

08 December 2022

Ashkan Shahbandi, Fang-Yen Chiu, … James G. Jackson

- Article
- Open Access
- Published: 05 December 2022

Dendritic cells direct circadian anti-tumour immune responses

- Chen Wang,
- Coline Barnoud,
- Mara Cenerenti,
- Mengzhu Sun,
- Irene Caffa,
- Burak Kizil,
- Ruben Bill,
- Yuanlong Liu,
- Robert Pick,
- Laure Garnier,
- Olga A. Gkountidi,
- Louise M. Ince,
- Stephan Holtkamp,
- Nadine Fournier,
- Olivier Michielin,
- Daniel E. Speiser,
- Stéphanie Hugues,
- Alessio Nencioni,
- Mikaël J. Pittet,
- Camilla Jandus &
- …
- Christoph Scheiermann

Show authors

Nature **volume 614**, pages 136–143 (2023)

- 17k Accesses
- 3 Citations
- 224 Altmetric

- Metrics details

Abstract

The process of cancer immunosurveillance is a mechanism of tumour suppression that can protect the host from cancer development throughout its lifetime[1,2]. However, it is unknown whether the effectiveness of cancer immunosurveillance fluctuates over a single day. Here we demonstrate that the initial time of day of tumour engraftment dictates the ensuing tumour size across mouse cancer models. Using immunodeficient mice as well as mice lacking lineage-specific circadian functions, we show that dendritic cells (DCs) and $CD8^+$ T cells exert circadian anti-tumour functions that control melanoma volume. Specifically, we find that rhythmic trafficking of DCs to the tumour draining lymph node governs a circadian response of tumour-antigen-specific $CD8^+$ T cells that is dependent on the circadian expression of the co-stimulatory molecule CD80. As a consequence, cancer immunotherapy is more effective when synchronized with DC functions, shows circadian outcomes in mice and suggests similar effects in humans. These data demonstrate that the circadian rhythms of anti-tumour immune components are not only critical for controlling tumour size but can also be of therapeutic relevance.

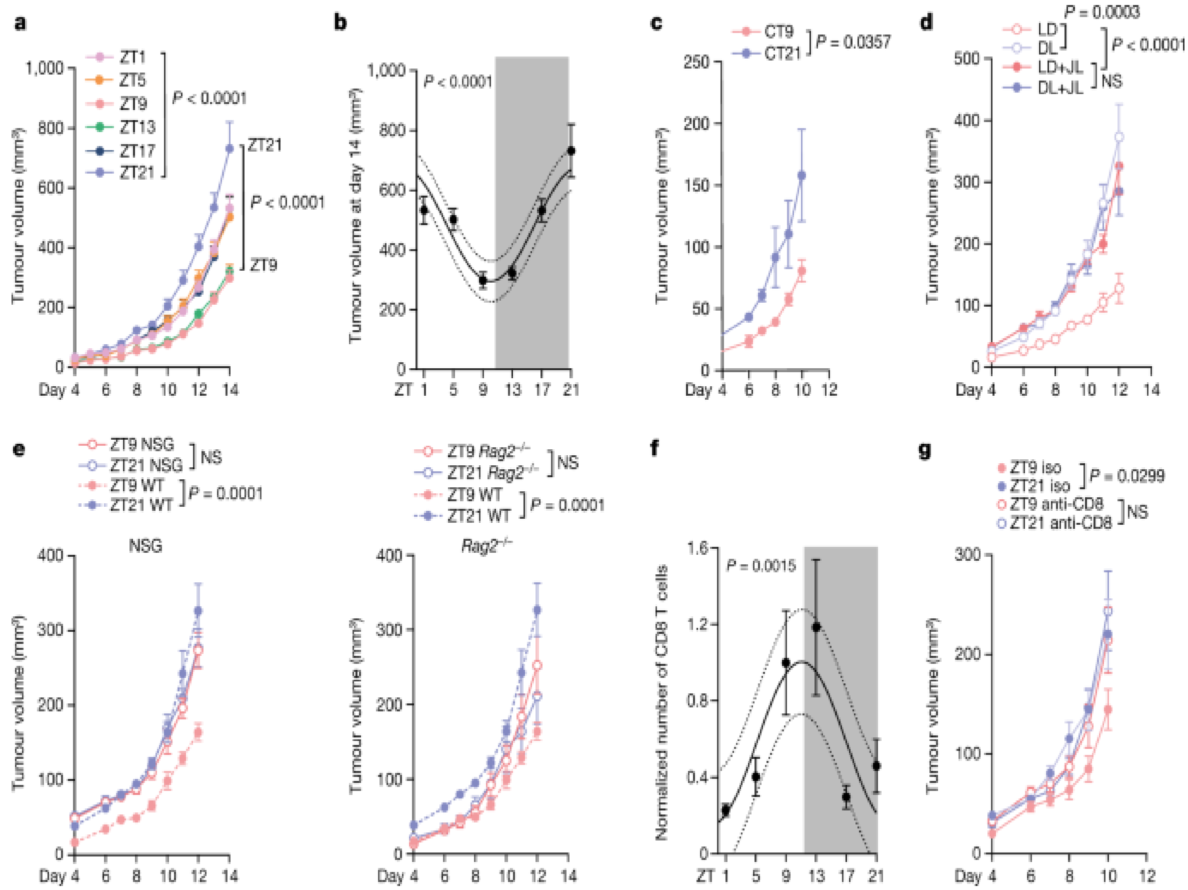
Main

The immune system provides sophisticated defence mechanisms that most often eliminate or contain the appearance of tumour cells in healthy tissue and prevent the development of life-threatening cancers[1,2]. Both the innate and adaptive arms of immunity show circadian (around 24 h) rhythmicity in their response[3,4,5,6,7,8,9,10], such that, even weeks after an initial stimulus is encountered, time-of-day immune effects are still observed[11,12,13,14,15,16]. There is evidence that cancer cells can exhibit a perturbation in their circadian clock components that drives cancer development[17]. However, the effect of a rhythmic immune system on tumour surveillance, and the effectiveness of treatments involving the immune system, remain unclear. Here we provide evidence that a circadian anti-tumour immune response controls tumour volume and the response to therapy.

Timed engraftment determines tumour size

To examine whether tumour volume depends on the time of day of tumour cell engraftment, we injected B16-F10 melanoma cells expressing ovalbumin (B16-F10-OVA) subcutaneously into cohorts of mice at six different times of day (that is, at zeitgeber time 1 (ZT1; 1 h after light onset in a 12 h–12 h light–dark environment, that is, morning), ZT5 (midday), ZT9 (afternoon), ZT13 (evening), ZT17 (midnight) and ZT21 (early morning)) and quantified tumour size over the ensuing two weeks. To control these data, animals were housed in distinct environmental light cabinets, 12 h phase-shifted to each other, enabling the simultaneous injection of the same batch of tumour cells into differently timed recipients. Tumour size was strongly affected by the time of day of engraftment, yielding significantly larger tumours when inoculated at late night (ZT21), and smaller tumours when inoculated in the late afternoon (ZT9–ZT13) (Fig. 1a,b and Extended Data Fig. 1a). We observed similar results in two orthotopic mammary carcinoma models (E0771 and 4T1) (Extended Data Fig. 1b–d) and a mouse colon carcinoma model (MC-38) (Extended Data Fig. 1e). This indicated that the time-of-day effect of engraftment on tumour size represented a phenomenon that is relevant across different tumour types and sites of engraftment. We further confirmed these data using quantitative imaging approaches with luciferase-expressing melanoma cells (B16-F10-OVA-Luc) (Extended Data Fig. 1f). Experiments using B16-F10 melanoma cell lines that did or did not express OVA yielded very similar results (Extended Data Fig. 1g). This indicated that the observed time-of-day effect of inoculation represented a general phenotype that is not affected by the potentially different immunogenicity of the tumour.

Fig. 1: The time of day of engraftment determines tumour size.



a, Tumour volume after engraftment of B16-F10-OVA cells at six different times of day. $n = 10$ mice from 2 independent experiments. Statistical analysis was performed using two-way analysis of variance (ANOVA). **b**, The tumour volume on day 14 from the experiments shown in **a**. Statistical analysis was performed using a cosinor analysis. **c**, Tumour volume after engraftment of B16-F10-OVA cells at two different times of day under constant darkness (DD) conditions. CT, circadian time. $n = 6$ mice from 2 independent experiments. Statistical analysis was performed using two-way ANOVA. **d**, Tumour volume after engraftment of B16-F10-OVA cells at two different times of day under light–dark (LD, $n = 6$ mice), inverted dark–light (DL, $n = 7$ mice) or jet lag (JL, $n = 7$ mice) conditions. For the jet lag condition, mice were placed into a 6 h or 12 h phase-delayed environment every 3 days. $n = 2$ independent experiments. Statistical analysis was performed using two-way ANOVA. **e**, Tumour volume after engraftment of B16-F10-OVA cells at two different times of day in NSG mice (left, $n = 10$ mice) or *Rag2*$^{-/-}$ mice (right). $n = 10$ (ZT9) and $n = 11$ (ZT21) mice. Control WT mice ($n = 9$) are plotted as a reference. Data are from two independent

experiments. Statistical analysis was performed using two-way ANOVA. **f**, Tumour-infiltrating $CD8^+$ T cells on day 14 from the experiment in **a**. From ZT1 to ZT21, $n = 10$, $n = 9$, $n = 10$, $n = 7$, $n = 10$, $n = 8$ mice from $n = 4$ independent experiments. Statistical analysis was performed using a cosinor analysis. **g**, Tumour volume after engraftment of B16-F10-OVA cells at two different times of day after anti-CD8 antibody depletion. Iso, isotype control. $n = 6$ mice from 2 independent experiments. Statistical analysis was performed using two-way ANOVA. The shaded areas indicate dark phases. For **a**–**g**, data are mean ± s.e.m. NS, not significant.

Source data

Circadian rhythms are defined by their persistence in the absence of environmental entraining cues, such as rhythmic light onset and offset. Transferring animals to complete darkness conditions did not alter the observed time-of-day-dependent differences, demonstrating that the effect is bona fide circadian in nature (Fig. 1c). However, switching mice to a 12 h inverted dark–light cycle inversed tumour size, demonstrating that the effect was not dependent on light per se but that it could be entrained by light, an additional feature of circadian rhythms (Fig. 1d and Extended Data Fig. 1h). The use of a mouse jet-lag protocol (Extended Data Fig. 1h) abrogated time-of-day differences and increased tumours, indicating that acutely altering lighting regimes negatively affected the disease outcome (Fig. 1d). Together, these data provide evidence that tumour size is highly governed by the initial time of day of engraftment, driven by circadian rhythms in the host that are entrained by light.

Circadian anti-tumour immune effects

To assess whether these effects were dependent on the immune system, we injected B16-F10-OVA melanoma cells during the day (ZT9) or at night (ZT21) into NSG mice, which lack both adaptive and innate immune cells. Importantly, the previously observed dependency of tumour volume on time-of-day engraftment was abrogated in these mice, indicating that the differences are mediated by the immune system (Fig. 1e). To define which arm of immunity was involved, we used *Rag2*$^{-/-}$ mice, which lack an adaptive immune system. Similar to NSG mice, time-of-day differences in

tumour size were ablated in these animals, demonstrating that the adaptive immune system is critical in mediating the phenotype (Fig. 1e).

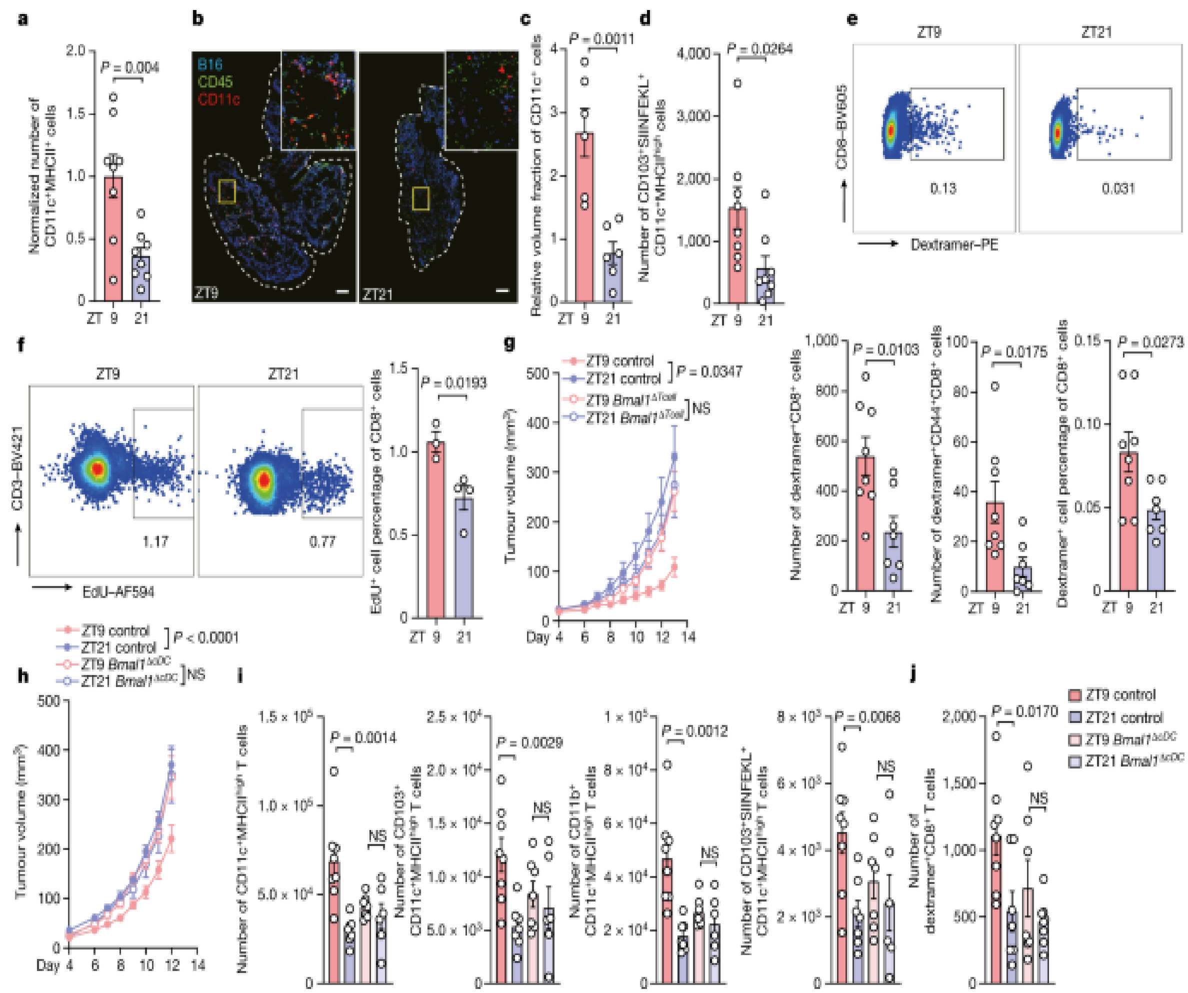
We next used flow cytometry to assess the immune cell infiltrates in tumours 14 days after inoculation. Tumours were collected at ZT1 to limit the variables to only the time of engraftment. The numbers of $CD8^+$ T cells were dependent on the time of engraftment, with cellularity peaking when tumour inoculation occurred during the day (ZT9) and troughing at night (ZT21) (Fig. 1f and Extended Data Fig. 2a). By contrast, the numbers of other leukocyte subsets were not affected (Extended Data Fig. 2b). To assess the functional relevance of rhythmicity in the tumour immune cell infiltrate, we used different antibodies to deplete specific subpopulations of leukocytes. Antibody-mediated depletion of $CD8^+$ T cells or $CD4^+$ T cells—but not of neutrophils—abrogated the time-of-day-dependent differences in tumour size (Fig. 1g and Extended Data Fig. 2c–j). However, only depletion of $CD8^+$ T cells increased the tumour volume (Fig. 1g); by contrast, depletion of $CD4^+$ T cells decreased the tumour volume (Extended Data Fig. 2c). These results indicate that $CD8^+$ T cells exert anti-tumorigenic effects in a time-of-day-dependent manner.

Rhythmic anti-tumour response in DCs

To investigate the mechanisms controlling the time-of-day-dependent effect of $CD8^+$ T cells on tumour size, we focused on the early events that potentially accounted for the observed effects. Using flow cytometry and quantitative imaging approaches to characterize the site of engraftment 4 h after tumour inoculation, we found that $CD11c^+MHCII^+$ cells represent the predominant leukocyte subset, with higher numbers of these cells when tumour cells were inoculated at ZT9 compared with ZT21 (Fig. 2a–c and Extended Data Fig. 3a–c). Furthermore, 24 h after tumour inoculation, we detected more leukocytes in the draining lymph nodes (dLN) of mice in which tumour cells were inoculated at ZT9 compared with at ZT21 (Extended Data Fig. 3d–e). Specifically, we observed more $CD4^+$ and $CD8^+$ T cells, including increased numbers of activated central memory ($CD44^+CD62L^+$) and naive ($CD44^-$ $CD62L^+$) T cells, after tumour cell engraftment at ZT9 (Extended Data Fig. 3e,f). These dLNs also contained

more $CD11c^+$ cells, including the $CD11b^+CD11c^+MHCII^{high}$, $CD103^+CD11c^+MHCII^{high}$, $CD11b^+CD11c^+MHCII^{low}$ and $CD8^+CD11c^+MHCII^{low}$ subsets (Extended Data Fig. 3g–i). This phenotype was also observed in the orthotopic mammary carcinoma as well as the colon carcinoma models (Extended Data Fig. 4a–g). Under sham conditions, time-of-day-dependent changes in the number of these cell types were observed but showed smaller differences (Extended Data Fig. 5a). To further identify relevant tumour-derived antigen-presenting cells (APCs) in the dLN, we used an antibody specific for the SIINFEKL peptide bound to H-$2K^b$ (Extended Data Fig. 3h,i). We found that APCs presenting this antigen predominantly displayed a $CD103^+CD11c^+MHCII^{high}$ phenotype. These $CD103^+CD11c^+MHCII^{high}(SIINFEKL–H\text{-}2K^b)^+$ cells were also more numerous in the dLNs of mice inoculated with tumour cells at ZT9 (Fig. 2d and Extended Data Fig. 3h,i). This phenotype was also observed in the orthotopic E0771-OVA breast cancer model (Extended Data Fig. 4c,d). No differences were observed in the processing of OVA antigen in $CD11c^+$ cells (Extended Data Fig. 5b), suggesting that changes in the levels of antigen presentation were not responsible for the phenotype.

Fig. 2: DCs respond rhythmically to tumour engraftment.



a, The number of cells at the skin engraftment site 4 h after engraftment of B16-F10-OVA cells at two different times of day. $n = 8$ mice from 2 independent experiments. Statistical analysis was performed using unpaired Student's *t*-tests. **b**,**c**, Imaging (**b**) and quantification (**c**) of $CD11c^{+}$ cells at the skin engraftment site 4 h after engraftment of B16-F10-OVA cells. $n = 6$ mice from 2 independent experiments. Statistical analysis was performed using unpaired Student's *t*-tests. For **b**, scale bars, 500 µm. **d**, The numbers of cells in the dLN 24 h after engraftment of B16-F10-OVA cells. $n = 8$ mice from 2 independent experiments. Statistical analysis was performed using unpaired Student's *t*-tests. **e**, $H\text{-}2K^{b}$–SIINFEKL dextramer staining of $CD8^{+}$ T cells in the dLN 72 h after engraftment of B16-F10-OVA cells. $n = 8$ (ZT9) and $n = 7$ (ZT21) mice from 2 independent experiments. Statistical analysis was performed using unpaired Student's *t*-tests. **f**, EdU staining gated on $CD3^{+}CD8^{+}$ T cells in the dLN 48 h after engraftment of B16-F10-OVA cells. $n = 3$ (ZT9) and $n = 4$ (ZT21) mice, representative of 2 independent

experiments. Statistical analysis was performed using unpaired Student's *t*-tests. **g**,**h**, Tumour volume after engraftment of B16-F10-OVA cells in *Cd4-cre:Bmal1^{flox}* mice (*n* = 8 (ZT9 control), *n* = 16 (ZT9 *cre*), *n* = 16 (ZT21 control) and *n* = 7 (ZT21 *cre*)) (**g**) and *Clec9a-cre:Bmal1^{flox}* mice (*n* = 17 (ZT9 control), *n* = 9 (ZT9 *cre*), *n* = 16 (ZT21 control) and *n* = 10 (ZT21 *cre*)) (**h**). *n* = 3 independent experiments. Statistical analysis was performed using two-way ANOVA. **i**, The cell numbers for the CD11c^{+}MHCII^{high} subsets in the dLN 24 h after engraftment of B16-F10-OVA cells in *Clec9a-cre:Bmal1^{flox}* mice. *n* = 8 (ZT9 control), *n* = 7 (ZT9 *cre*), *n* = 7 (ZT21 control) and *n* = 6 (ZT21 *cre*) from 2 independent experiments. Statistical analysis was performed using unpaired Student's *t*-tests. **j**, H-2K^{b}–SIINFEKL dextramer staining of CD8^{+} T cells in the dLN 72 h after engraftment of B16-F10-OVA cells in *Clec9a-cre:Bmal1^{flox}* mice. *n* = 8 (ZT9 control), *n* = 7 (ZT9 *cre*), *n* = 7 (ZT21 control) and *n* = 8 (ZT21 *cre*) mice from 2 independent experiments. Statistical analysis was performed using unpaired Student's *t*-tests. For **a**, **c**, **d** and **f**–**j**, data are mean ± s.e.m. All *t*-tests were two-tailed.

Source data

Using dextramer staining to detect endogenous T cells specific for H-2K^{b}–SIINFEKL in the dLN 72 h after tumour engraftment, we detected higher numbers and proportions of antigen-specific CD44^{+}CD8^{+} T cells when tumour inoculation was performed at ZT9 compared with ZT21 in both the melanoma and mammary carcinoma models (Fig. 2e and Extended Data Fig. 4e). We also observed significantly more EdU^{+}CD8^{+} T cells and EdU^{+}CD4^{+} T cells in the dLN 48 h after tumour inoculation at ZT9 compared with ZT21, demonstrating a higher T cell proliferation rate after ZT9 tumour inoculation (Fig. 2f and Extended Data Fig. 5c). In an analogous manner, we performed dextramer staining to detect endogenous T cells specific for a peptide of the neoantigen Adpgk (H-2D^{b}/ASMTNMELM), expressed by MC-38 colon carcinoma cells. We detected higher numbers of Adpgk^{+}-neoantigen-specific CD8^{+} T cells in this tumour model in the dLN 72 h after tumour engraftment when tumour inoculation was performed at ZT9 compared with ZT21 (Extended Data Fig. 4f,g). These data indicate that tumour-antigen-specific CD8^{+} T cells are generated in a time-of-day-

dependent manner in the dLN against OVA antigen as well as neoantigen, and similar mechanisms are involved in the subcutaneous and orthotopic engraftment setting.

Contribution of DC and T cell clocks

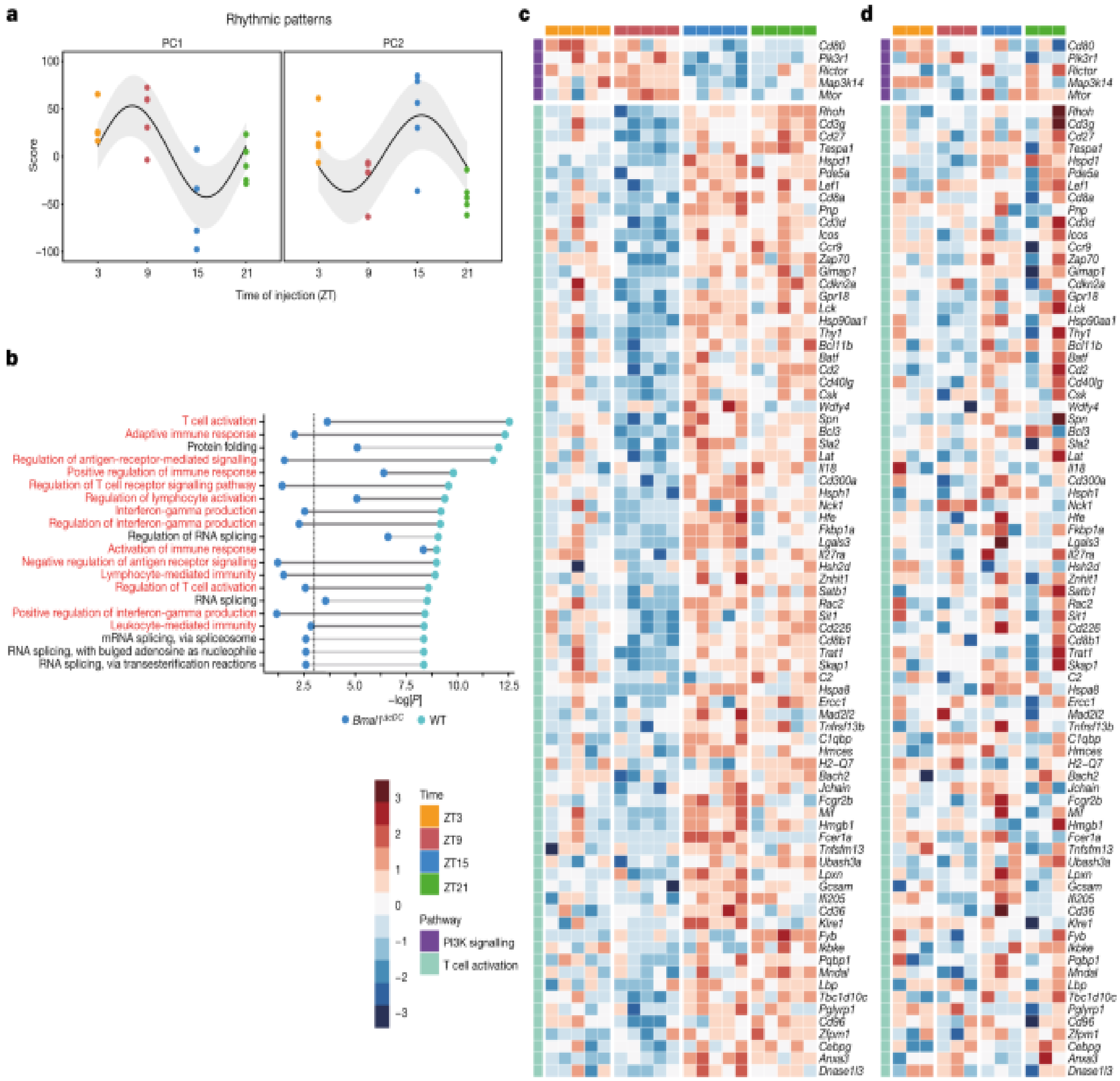
To investigate whether the differences were driven by immune-cell-intrinsic mechanisms, we used *Cd4-cre:Bmal1^{flox}* mice, in which the key circadian clock gene *Arntl* (also known as *Bmal1*) is specifically deleted in T cells (*Bmal1$^{\Delta Tcell}$*), rendering them arrhythmic. *Bmal1$^{\Delta Tcell}$* mice showed similar kinetics of tumour volume when tumour cells were inoculated either at ZT9 or ZT21, demonstrating the importance of T-cell-intrinsic rhythms in the control of tumour size (Fig. 2g). Similarly, *Clec9a-cre:Bmal1^{flox}* mice, which lack *Bmal1* expression in conventional DCs (*Bmal1$^{\Delta cDC}$*), showed comparable tumour size kinetics when tumour cells were inoculated either at ZT9 or ZT21 (Fig. 2h). This demonstrated that BMAL1 and cell-autonomous circadian oscillations in both DCs and T cells are critical for the time-of-day differences in tumour volume. Mechanistically, *Bmal1* deletion in cDCs abrogated the differences in total and antigen-specific DC numbers in the dLN after tumour engraftment, in contrast to control mice (Fig. 2i). Furthermore, dextramer staining in *Bmal1$^{\Delta cDC}$* mice revealed reduced antigen-specific $CD8^+$ T cell levels and abrogated time-of-day-dependent differences (Fig. 2j). These data demonstrate that DC and T cell autonomous circadian clocks are responsible for the time-of-day-dependent anti-tumour effects, with DCs governing rhythmic $CD8^+$ T cell responses.

DCs govern rhythmic anti-tumour immunity

To obtain global information on DC changes after tumour inoculation at different times of the day, we performed RNA-sequencing (RNA-seq) analyses of the subset of $CD11c^+MHCII^{high}$ migratory DCs in the dLN collected 24 h after tumour engraftment or sham conditions, inoculated at ZT3, ZT9, ZT15 or ZT21. We observed strong time-of-day-dependent differences in overall gene expression, indicating differences in DC functionality. First, we found that $CD11c^+MHCII^{high}$ cells exhibited rhythmicity in the expression of clock genes and clock-controlled genes

(Extended Data Fig. 6a); the expression of these genes was sufficient to define the time of day from which the cells were derived (Extended Data Fig. 6b). Next, we detected two main clusters of oscillatory genes, one of which was expressed more highly in the morning (ZT9) and the other in the evening (ZT15) (Fig. 3a,b). Whereas the morning cluster consisted mainly of metabolic genes—with the exception of the co-stimulatory molecule CD80—the second cluster was highly enriched in T cell activation pathways (Figs. 3b–d and 4a and Extended Data Fig. 6c,d). By contrast, RNA-seq analyses of $CD11c^{+}MHCII^{high}$ cells collected from *Bmal1*$^{\Delta cDC}$ mice showed altered rhythmicity, gene expression patterns and cellular pathways compared with the control mice (Figs. 3b,d and 4a and Extended Data Figs. 6c and 7a–c). These data indicate that, in addition to differences in DC numbers, the rhythmicity in DC co-stimulatory factors that was specific for the tumour scenario (Extended Data Fig. 8a–c) could be responsible for the generation of rhythmic $CD8^{+}$ T cell activation phenotypes.

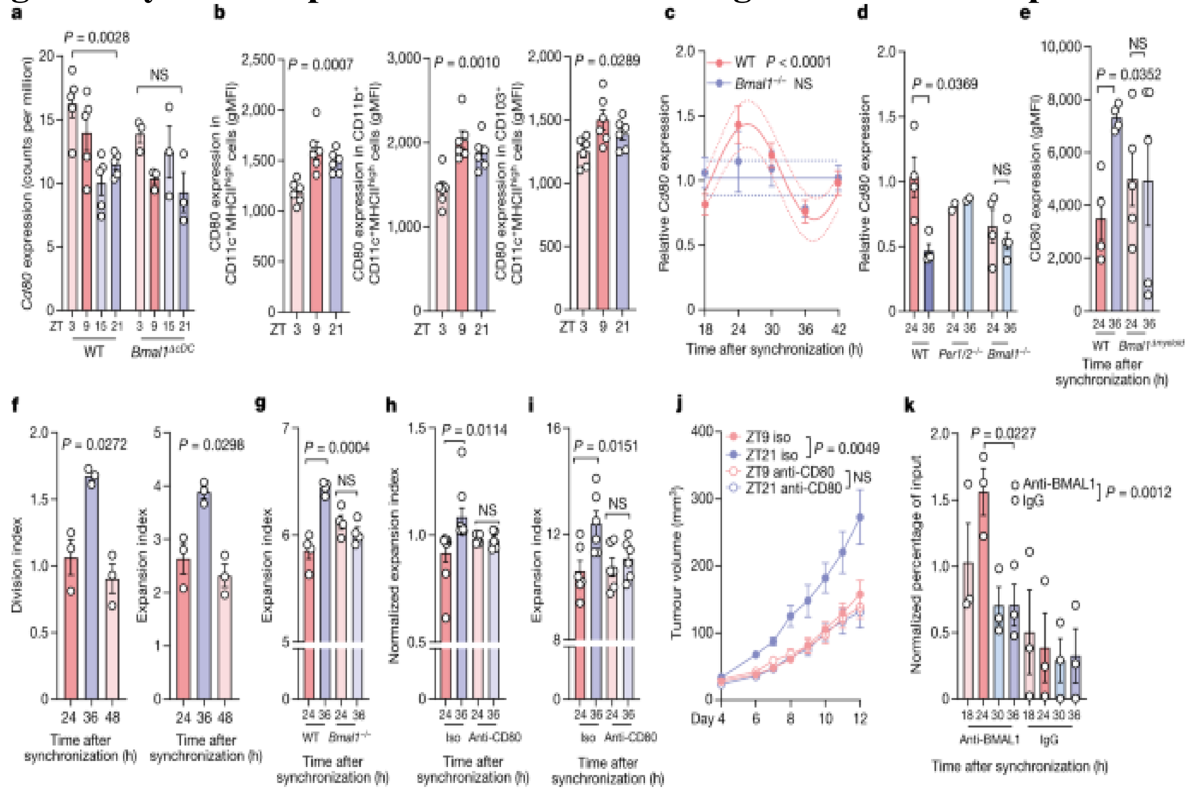
Fig. 3: DCs exhibit circadian gene expression patterns.



a–**d**, RNA-seq analysis of CD11c^{+}MHCII^{high} cells in the dLN 24 h after engraftment of B16-F10-OVA cells at ZT3, ZT9, ZT15 or ZT21 in control mice ($n = 5$ mice) or *Clec9a-cre:Bmal1^{flox}* mice ($n = 3$ mice). $n = 2$ independent experiments. **a**, Principal component (PC) analyses of the two main peaks in gene expression oscillation in control mice. $n = 5$ mice. Statistical analysis was performed using a cosinor analysis. **b**, Significantly enriched Gene Ontology pathways from PC2 in the control cells shown in **a**, with T cell activation pathways highlighted in red, shown for control and *Clec9a-cre:Bmal1^{flox}* CD11c^{+}MHCII^{high} cells. The vertical dashed line represents the significant *P* values, which were calculated using hypergeometric tests. **c**,**d**, Significantly expressed genes in the CD28-

dependent PI3K–AKT signalling pathway (top) or T cell activation pathways (bottom) in control mice (**c**) and the lack of significance in *Clec9a-cre:Bmal1*^{flox} mice (**d**).

Fig. 4: Rhythmic expression of CD80 in DCs governs T cell responses.



a, Expression (counts per million (CPM)) of *Cd80* in CD11c^{+}MHCII^{high} cells from control ($n = 5$) or *Clec9a-cre:Bmal1*^{flox} ($n = 3$) mice. Statistical analysis was performed using one-way ANOVA. **b**, CD80 expression in DCs subsets was determined using flow cytometry in the dLN 24 h after engraftment of B16-F10-OVA cells. $n = 6$ mice from 2 independent experiments. Statistical analysis was performed using one-way ANOVA. gMFI, geometric mean fluorescence intensity. **c**, *Cd80* mRNA expression after synchronization of LPS-matured BMDCs from WT ($n = 10$) and *Bmal1*$^{-/-}$ ($n = 4$) mice from 2 independent experiments. Statistical analysis was performed using a cosinor analysis. **d**, *Cd80* mRNA expression after synchronization of BMDCs generated from WT ($n = 4$) *Per1*$^{-/-}$*Per2*$^{-/-}$ ($n = 2$) or *Bmal1*$^{-/-}$ ($n = 4$) mice. $n = 2$ independent experiments. Statistical analysis was performed using unpaired Student's *t*-tests. **e**, Floy cytometry analysis of CD80 protein expression in synchronized *Lyz2-cre:Bmal1*^{flox}

BMDCs. $n = 4$ (control) and $n = 5$ (*cre*) mice from 2 independent experiments. Statistical analysis was performed using paired Student's *t*-tests. **f**–**h**, In vitro co-culture proliferation experiments with OT-I $CD8^+$ T cells and SIINFEKL-loaded BMDCs generated from WT mice ($n = 3$ mice from 2 independent experiments) (**f**) or *Bmal1*$^{-/-}$ mice ($n = 4$, from 2 independent experiments) (**g**) or after anti-CD80 antibody treatment (n = 9 (control) and $n = 5$ (anti-CD80) mice from 9 independent experiments) (**h**). Statistical analysis was performed using one-way ANOVA (**f**), an unpaired Student's *t*-test (**g**) and a paired Student's *t*-test (**h**). **i**, In vitro co-culture proliferation experiments with naive $CD8^+$ T cells treated with an anti-CD3 antibody and WT BMDCs in the presence of absence of an anti-CD80 antibody. $n = 3$ mice, 2 replicates each, from 2 independent experiments. Statistical analysis was performed using unpaired Student's *t*-tests. **j**, The tumour volume after engraftment of B16-F10-OVA cells at two different times of day, and after treatment with an anti-CD80 antibody or isotype control. $n = 10$ mice from 2 independent experiments. Statistical analysis was performed using two-way ANOVA. **k**, ChIP analysis of BMAL1 binding to the promoter region of *Cd80* in synchronized BMDCs. $n = 3$ mice from 2 independent experiments. Statistical analysis was performed using unpaired Student's *t*-tests. For **a**–**k**, all data are mean ± s.e.m. All *t*-tests were two-tailed.

Source data

CD80 expression was confirmed to be time-of-day dependent in different $CD11c^+$ subsets at the protein level on the basis of flow cytometry analysis (Fig. 4b). To investigate whether oscillations in CD80 were driven by a cell-autonomous circadian rhythm—independent of the environment as indicated by the RNA-seq data (Fig. 4a)—we performed in vitro synchronization assays of immature as well as LPS-matured bone-marrow-derived DCs (BMDCs) using a serum shock[18,19] (Fig. 4c and Extended Data Fig. 9a–c). We detected significant differences in *Cd80* expression at different timepoints after BMDC synchronization (Fig. 4c and Extended Data Fig. 9c). By contrast, we observed no circadian differences in the expression of various cytokines but, rather, a timer-dependent progressive change through time (Extended Data Fig. 9d). Furthermore, circadian differences in *Cd80* expression were abrogated in $CD11c^+MHCII^{high}$ cells collected from

Bmal1$^{\Delta cDC}$ in vivo as well as BMDCs generated from mice lacking overall expression of circadian genes in vitro (*Bmal1*$^{-/-}$ or *Per1*$^{-/-}$*Per2*$^{-/-}$) (Fig. 4a,c,d). To further test whether circadian rhythmicity in myeloid cells was necessary to control *Cd80* expression, we assessed BMDCs produced from mice lacking BMAL1 specifically in myeloid cells (*Lyz2-cre:Bmal1*^{flox};*Bmal1*$^{\Delta myeloid}$); again, we found that the loss of BMAL1 abrogated time-of-day differences in CD80 expression (Fig. 4e). Taken together, these findings indicate a critical role for the circadian clock machinery in controlling CD80 expression.

To investigate whether the circadian rhythmicity in DC gene expression patterns had a functional and causal consequence on T cell activation, we performed co-culture experiments of synchronized, SIINFEKL-pulsed BMDCs with non-synchronized OVA-specific OT-I CD8$^+$ T cells. This approach enabled us to observe that the rhythmicity of DCs directly controlled that of T cells. Indeed, the proliferation of OT-I CD8$^+$ T cells was highly dependent on the rhythmic phase in which the DCs were located (Fig. 4f). BMDCs collected 36 h after synchronization induced stronger OT-I T cell proliferation compared with BMDCs collected 24 h after synchronization (Fig. 4f). By contrast, BMAL1-deficient BMDCs did not induce rhythmic OT-I CD8$^+$ T cell proliferation (Fig. 4g). Furthermore, treatment with an anti-CD80 antibody abrogated time-of-day-dependent differences in OT-I CD8$^+$ T cell proliferation (Fig. 4h), demonstrating the relevance of CD80 in the rhythmic CD8$^+$ T cell response.

To specifically investigate the importance of rhythmicity in co-stimulatory signals provided by BMDCs, we performed co-culture experiments of synchronized, SIINFEKL-pulsed BMDCs with non-synchronized OVA-specific OT-I CD8$^+$ T cells, as before, but bypassing MHCI–TCR interactions using an anti-CD3 antibody. Co-stimulatory signals, in the presence of isotype-matched control antibodies, were sufficient to promote time-of-day-dependent differences in OT-I CD8$^+$ T cell proliferation; however, treatment with an anti-CD80 antibody abrogated these time-of-day-dependent differences (Fig. 4i). Finally, antibody treatment against CD80 abrogated the differences in tumour volume after time-of-day-dependent engraftment (Fig. 4j), demonstrating the functional relevance of

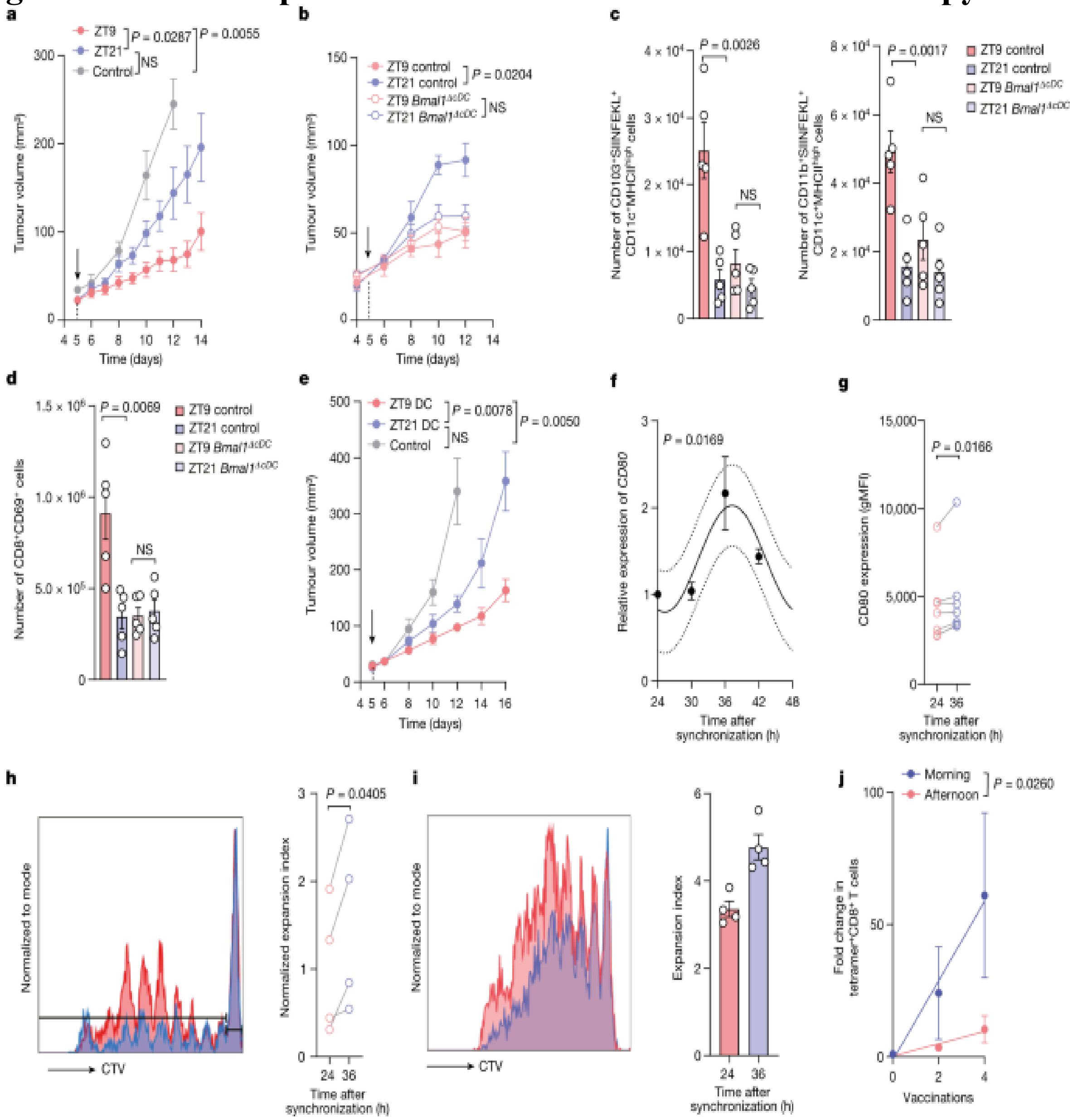
CD80 in vivo in driving differences in tumour size. These data indicate that CD80 is a critical molecule in this process.

Moreover, we identified the presence of BMAL1-binding sites, namely canonical enhancer boxes (E-boxes), in the promoter region of the *Cd80* gene. This suggests that *Cd80* expression is directly regulated by the circadian clock (Extended Data Fig. 9e). Indeed, using chromatin immunoprecipitation (ChIP) assays, we confirmed rhythmic binding of BMAL1 to the *Cd80* promoter (Fig. 4k and Extended Data Fig. 9f). Together, these data demonstrate that circadian rhythms in DCs direct the rhythms of T cell proliferation, a phenomenon that is dependent on the rhythmic expression of CD80, which is under direct transcriptional control of the clock gene *Bmal1*.

Vaccination as tumour chrono-immunotherapy

To evaluate the translational potential of our findings, we examined tumour chrono-immunotherapy. Specifically, we studied mice that were inoculated with B16-F10-OVA melanoma cells at ZT9 and then immunized with OVA either during the day (ZT9) or at night (ZT21). This setting limited the time-of-day information to the timepoint of vaccination only. Notably, we found that tumour volume was strongly suppressed by the vaccine when administered to wild-type (WT) mice at ZT9 compared with ZT21 (Fig. 5a), even when the relative incubation time of the vaccine was substantially longer for ZT21 than for ZT9 (Extended Data Fig. 10a). By contrast, $Bmal1^{\Delta cDC}$ mice vaccinated with OVA at ZT9 or ZT21 showed similar tumour sizes (Fig. 5b). An analysis of dLNs from WT mice 24 h after vaccination revealed a higher number of SIINFEKL-presenting DCs in mice that were vaccinated at ZT9 compared with those vaccinated at ZT21 (Fig. 5c and Extended Data Fig. 10b). Furthermore, this increase at ZT9 coincided with higher numbers of $CD69^+CD8^+$ and $CD69^+CD4^+$ T cells (Fig. 5d and Extended Data Fig. 10c). By contrast, DC and T cell numbers and phenotypes remained similar in $Bmal1^{\Delta cDC}$ mice vaccinated at ZT9 or ZT21 (Fig. 5c,d and Extended Data Fig. 10b,c). These data indicate a key role of cDC rhythmicity in generating a productive anti-tumour immune response after treatment.

Fig. 5: Chronotherapeutic vaccination as tumour immunotherapy.



a, Tumour volume after engraftment of B16-F10-OVA cells at ZT9 and OVA vaccination on day 5 (arrow) at ZT9 or ZT21. $n = 12$ (vaccinated) and $n = 3$ (unvaccinated control) mice from 2 independent experiments. Statistical analysis was performed using two-way ANOVA. **b**, Tumour volume after engraftment of B16-F10-OVA cells with OVA vaccination on day 5 (arrow) at ZT9 or ZT21 in control or *Clec9a-cre:Bmal1*^{flox} mice. $n = 5$ mice from 2 independent experiments. Statistical analysis was performed using two-way ANOVA. **c**,**d**, The cell numbers for DC subsets (**c**) and T cells (**d**) in the dLN 24 h after OVA vaccination (on day 5 after B16-F10-

OVA cell engraftment) in control or *Clec9a-cre:Bmal1^{flox}* mice. $n = 5$ mice from 2 independent experiments. Statistical analysis was performed using unpaired Student's *t*-tests. **e**, The tumour volume after engraftment of B16-F10-OVA cells with SIINFEKL-loaded BMDC vaccination on day 5 (arrow) at ZT9 or ZT21. $n = 6$ mice from 2 independent experiments. Statistical analysis was performed using two-way ANOVA. **f**, mRNA expression of *CD80* in hMoDCs after synchronization. $n = 3$ patients. Statistical analysis was performed using a cosinor analysis. **g**, Human CD80 protein expression was analysed using flow cytometry in hMoDCs after synchronization. $n = 7$ patients. Statistical analysis was performed using paired Student's *t*-tests. **h**, In vitro co-culture proliferation experiments with human naive $CD8^+$ T cells and synchronized hMoDCs. $n = 4$ patients. Statistical analysis was performed using paired Student's *t*-tests. **i**, In vitro co-culture proliferation experiments with antigen-specific $CD8^+$ T cells from patients with melanoma and synchronized HLA-$A2^+$ hMoDCs. Data are technical replicates, representative of two donors from two independent experiments. Statistical analysis was performed using unpaired Student's *t*-tests. **j**, Fold change in Melan-A-specific T cells after 2 and 4 vaccinations (with Melan-A peptide, CpG 7909 and incomplete Freund's adjuvant) in the morning ($n = 6$ patients) or afternoon ($n = 4$ patients). Statistical analysis was performed using a linear regression analysis. For **a**–**g**, **i** and **j**, data are mean ± s.e.m. All *t*-tests were two-tailed.

Source data

We further performed vaccination experiments at ZT9 and ZT21 in a scenario in which tumours were also inoculated at ZT9 and ZT21, therefore assessing the contribution of time-of-day effects in both the timing of tumour inoculation and the timing of vaccination. In these experiments, we observed that the timing of vaccination had a greater impact than the timing of tumour inoculation on tumour burden (Extended Data Fig. 10d). We also confirmed the time-of-day-dependent differences in vaccine efficacy in additional experiments in which two vaccinations were performed at ZT9 or ZT21 several days apart (Extended Data Fig. 10e). These data demonstrate that the timing of vaccination is a powerful means of reducing tumour size, and that rhythmicity of cDCs has a critical role in this process.

To bypass any endogenous DCs acting as potential APCs in this scenario, we performed vaccinations with subcutaneous injections of SIINFEKL-peptide-loaded BMDCs during the day (ZT9) or at night (ZT21). These experiments showed very similar results to the antigen vaccination studies, with suppressed tumour volume after daytime administration of the BMDCs (Fig. 5e). To assess whether these observations could translate to humans, we generated human monocyte-derived DCs (hMoDCs) from $CD14^+$ primary monocytes isolated from buffy coats from healthy donors and synchronized them in vitro. hMoDCs exhibited circadian expression of *CD80* as well as of the clock gene *PER2* (Fig. 5f,g and Extended Data Fig. 10f), analogous to the results obtained in mice. Furthermore, co-culture experiments using synchronized hMoDCs together with naive $CD8^+$ T cells isolated from the same healthy donors and stimulated with anti-human CD3 antibody—therefore limiting rhythmicity to co-stimulatory factors in DCs only—showed increased T cell proliferation in hMoDCs 36 h after synchronization compared with the 24 h timepoint (Fig. 5h).

Moreover, by generating HLA-$A2^+$ MDCs that were pulsed with Melan-$A_{26–35(A27L)}$ peptide (ELAGIGILTV; hereafter, ELA) and co-cultured with HLA-A2/ELA-specific $CD8^+$ T cell clones derived from patients with malignant melanoma[20], we observed time-of-day-dependent differences in the T-cell-proliferation capacity (Fig. 5i). This indicated that the rhythmic anti-tumour responses that we observed in mice were also present in human cells. Indeed, using retrospective time-of-day analyses of a tumour vaccination trial including 10 HLA-$A2^+$ patients with advanced malignant melanoma[20], we observed that time-of-day-dependent differences in vaccine administration result in increased Melan-A-specific $CD8^+$ T cells in the patients' blood, when vaccinations were performed in the morning compared with during the afternoon (Fig. 5j). Together, these data provide evidence for a role of time of day in tumour engraftment and in the efficacy of cancer immunotherapy in mice and humans.

Here we focused mostly on a mouse model of melanoma, while our additional data indicate that other cancer types are also affected by a rhythmic immune system; however, whether similar immune mechanisms are involved in other tumour models remains to be formally demonstrated. Furthermore, our initial patient data indicate the importance of considering

the time of day for the administration of cancer immunotherapy. By extension, it is possible that the time of day for the administration of any other treatment that involves activation of the immune system may matter. Given the relative simplicity of controlling this timing parameter in the clinic, it seems important to conduct prospective clinical trials that include sufficient numbers of patients and that can test whether the timing of injection of a given treatment improves the anti-tumour response and the patient's clinical outcome.

Methods

Animals

C57BL/6N and NSG mice were purchased from Charles River. BALB/c mice were purchased from Envigo. *Rag2*$^{-/-}$ mice (gift from W. Reith) were bred at Charles River. Other transgenic mouse lines were bred at ENVIGO: *Bmal1*$^{flox/flox}$, *Cd4-cre* (both purchased from Jackson Labs) and *Clec9a-cre* (gift from B. Schraml). Transgenic mice were maintained as homozygous for *Bmal1*$^{flox/flox}$ and heterozygous for the relevant *cre*. CD45.1 OTI (gift from W. Reith) mice and *Bmal1*$^{-/-}$ (gift from C. Dibner) mice were bred in house. All mice used were females at 8–12 weeks of age. Mice were housed under a 12 h–12 h light–dark schedule with food and water ad libitum. When multiple timepoints were investigated simultaneously, light-tight cabinets (Techniplast) were used to shift animals to the respective phase for a minimum of 1 day per 1 h of shift before the experiments. Treatment times correspond to Zeitgeber time (ZT) and indicate the timing relative to lights on in the animal facility such that ZT1 is 1 h after lights on (morning), ZT7 is 7 h after lights on (day time), ZT13 is 1 h after lights off (evening) and ZT19 is 7 h after lights off (night time). Animals were humanely euthanized if the tumour diameter reached 1.5 cm. All of the animal procedures and experiments were approved and performed in accordance with the guidelines of the animal research committee of Geneva, Switzerland or by the Italian Istituto Superiore di Sanità (ISS).

Tumour cell lines and inoculation

B16-F10 (ATCC) and B16-F10-OVA melanoma cells (from S. Hugues laboratory) were maintained in RPMI (Gibco) supplemented with 10% heat-inactivated FCS (Gibco), 100 μmol l^{-1} penicillin–streptomycin (Gibco) and 50 mmol l^{-1} β-mercaptoethanol (Gibco). B16-F10-OVA-Luc cells were created using ready-to-use lentivirus (GenTarget, LVP324) according to the manufacturer's instructions. Transduced cells were selected by puromycin (Thermo Fisher Scientific, A1113803) and isolated by fluorescence-activated cell sorting (FACS). MC38 mouse colon adenocarcinoma cells (from S. Hugues laboratory) were maintained in DMEM (Gibco), 10% heat-inactivated FCS, 100 μmol l^{-1} penicillin–streptomycin and 50 mmol l^{-1} β-mercaptoethanol. E0771 and E0771-OVA cells (from S. Hugues laboratory) were maintained in RPMI (Gibco) supplemented with 10% heat-inactivated FCS (Gibco), 100 μmol l^{-1} penicillin–streptomycin (Gibco) and 50 mmol l^{-1} β-mercaptoethanol (Gibco). The 4T1 cell line was purchased from ATCC and maintained in RPMI1640 medium supplemented with 10% heat-inactivated FBS, penicillin (50 U ml^{-1}) and streptomycin (50 μg ml^{-1}) (LifeTechnologies). Cell lines were used by passage 10 and tested negative for Mycoplasma. Unless otherwise specified, 5×10^5 tumour cells in 100 μl PBS were injected subcutaneously into the right flank of mice, under isoflurane anaesthesia. A total of 5×10^5 4T1, E0771, or E0771-OVA cells resuspended in PBS were injected orthotopically into the fourth abdominal fat pad of BALB/c (4T1) or C57BL/6 (E0771, E0771-OVA) female mice under ketamine–xylazine anaesthesia. Tumour volume was monitored every 1 to 2 days using callipers and calculated by length × width × width/2. The time of day of measurements did not influence tumour volume (data not shown). In a sham-injection experiment, 100 μl of PBS was injected subcutaneously without tumour cell injection.

IVIS imaging

d-Luciferin (Abcam, ab143655) was injected intraperitoneally into mice at a dose of 75 mg per kg body weight. Mice were anaesthetized with isoflurane and placed into the abdominal position. Images were collected 8 min after luciferin injection using the IVIS Imaging System (Xenogen), and photons emitted from the tumour were quantified using Living Image Software (Xenogen).

BMDCs

BMDCs were cultured as previously described[21], with complete culture medium (RPMI, 10% heat-inactivated FCS, 2 mM l-glutamine, 1% penicillin–streptomycin, 50 μM β-mercaptoethanol) supplemented with 20 ng ml^{-1} recombinant mouse GM-CSF (Peprotech). The medium was refreshed every 3 days. At day 10, all non-adherent and semi-adherent cells were collected in complete medium supplemented with 10 ng ml^{-1} GM-CSF and stimulated with 100 ng ml^{-1} lipopolysaccharide (LPS, L4516, Sigma-Aldrich) for 24 h.

BMDC synchronization

Cells were synchronized as previously described[18]. In brief, an equal volume of horse serum (Sigma-Aldrich, h1270) was prewarmed and added directly to the dish (serum shock). After incubation for 2 h at 37 °C under 5% CO_2, cells were washed and resuspended in complete medium.

Tissue digestion and single-cell preparation

The draining inguinal lymph node was collected and chopped into small pieces, then digested in 1 ml RPMI containing 1 mg ml^{-1} collagenase IV (Worthington Biochemical), 40 μg ml^{-1} DNase I (Roche, 04716728001) and 2% heat-inactivated FCS for 15 min at 37 °C using a thermoblock. Skin tissue was digested in RPMI containing 1 mg ml^{-1} collagenase IV, 2 mg ml^{-1} Dispase II (Roche), 40 μg ml^{-1} DNase I and 2% heat-inactivated FCS for 30 min at 37 °C. Chopped tumour tissue was digested using 1 mg ml^{-1} collagenase IV, 40 μg ml^{-1} DNase I and 2% heat-inactivated FBS for 30 min at 37 °C, and the remaining tumour went through 30 min further digestion using 1 mg ml^{-1} collagenase D (Roche). Cells were rinsed through a 70 μm cell strainer to obtain single-cell suspensions.

Flow cytometry

Single-cell suspensions were prepared and incubated with mouse or human Fc receptor block (anti-mouse CD16/32, BioLegend, human FcR blocking

reagent, Miltenyi Biotec) for 10 min at room temperature. After incubation, unless specified otherwise, the antibody mix was added directly into the cell suspension and incubated for 15 min at 4 °C.

The following anti-mouse antibodies were used for immunostaining: CD45 (30-F11, BUV395, BUV737, BD, 564279, 748371; FITC, BioLegend, 103107), CD45.1 (A20, PE, BioLegend, 110707), CD3e (145-2C11, BUV395, BD563565, APC, BioLegend, 100312; KT3.1.1, BV421, BioLegend, 155617), CD4 (GK1.5, BV650, BD, 563232), CD8a (53-6.7, BV605, BD, 563152; APC, BioLegend, 100711), CD11c (HL3, BUV737, BD612796, N418, PE, BioLegend, 117307), CD19 (1D3, BB700, BD, 566412), CD86 (GL1, BUV395, BD, 564199), CD80 (16-10A1, PE/Cy5, BioLegend, 104711), CD103 (2E7, BV421, BioLegend, 121421), NK1.1 (PK136, PE/Cy5, BioLegend, 108715), MHCII (M5/114.15.2, BV421, BV711, BV650, BioLegend, 107631, 107643, 107641), CD40 (1C10, PerCP-eFluor710, eBioscience, 46-0401-82), CD69 (H1.2F3, BUV737, BD, 612793; BV421, BioLegend, 104527), Ly6G (1A8, BV785, BioLegend, 127645), Ly6C (HK1.4, AF700, BioLegend, 128023). The following anti-human antibodies were used for immunostaining: HLA-DR (G46-6, BV480, BD, 566154), CD11C (B-ly6, BV711, BD, 563130), CD45RA (HI100, PE, BD, 555489), CD25 (2A3, BUV737, BD, 612807), CD44 (G44-26, APC/H7, BD, 560532), CD62L (DREG-56, BV510, BD, 563203), CD8 (RPA-T8, BUV395, BD, 563795), CCR7 (G043H7, BV785, BioLegend, 353230), CD3 (BW264/56, APC, Miltenyi Biotec, 130-113-687).

For peptide–MHC-dextramer staining, 10 µl dextramer (PE-H-2Kb, SIINFEKL, or APC-H-2Db Adpgk, Immudex) were added and incubated at room temperature for 15 min. Anti-mouse H-2K^b bound to SIINFEKL antibody staining (25-D1.16, APC, PE/Cy7, BioLegend, 141605) was performed at 37 °C for 15 min. Cells were washed and resuspended in 300 µl FACs buffer with viability dye (DAPI, BioLegend, 3 µM; propidium iodide, Invitrogen, 1.7 µg ml^{-1}; or DRAQ7, BioLegend, 2 µM) and characterized using an 18-colour BD LSR Fortessa (BD Biosciences) system or Beckman Coulter Cytoflex. Acquired data were analysed using FACSDiva 6 (BD Biosciences) and FlowJo 10 (BD). Cell counts were calculated using Counting Beads (C36950, C36995, Thermo Fisher Scientific).

When intracellular staining needed to be performed, cells were first stained with viability dye (eBioscience Fixable Viability Dye eFluor 780, 65-0865-18), followed by surface staining as previously described. For intracellular staining, cells were fixed and permeabilized using the FOXP3/Transcription Factor Staining Buffer Set (eBioscience, 00-5523-00). After washing with permeabilization buffer, the intracellular antibody (anti-mouse FOXP3, MF-14, AF647, BioLegend, 126408) was added and incubated for 30 min at room temperature.

RNA extraction, reverse transcription and qPCR

Cells were collected at the indicated timepoints and lysed using Trizol Reagent (Invitrogen). Tissues were homogenized in Trizol (Invitrogen) using a Precellys 24 (Bertin) bead mill homogenizer. Lysed and homogenized samples were processed using the Direct-zol RNA MiniPrep Kit (Zymo Research) according to the manufacturer's instructions. RNA quantity and quality was analysed using the Nanodrop 2000 (Thermo Fisher Scientific) or Bioanalyzer. Reverse transcription was performed using the PrimeScript RT Reagent Kit (Takara) according to the manufacturer's instructions. Quantitative PCR (qPCR) analyses were performed using PowerUp SYBR Green (Applied Biosystems); a list of the primer sequences is provided in Supplementary Table 4. Quantification of the transcript was performed using the $2^{-\Delta\Delta Ct}$ method using *Rplp0*, *Rpl32* and/or *Gapdh* as internal reference genes.

In vivo antibody treatments

To deplete specific leukocyte subsets, depletion antibodies or isotype control were injected intraperitoneally 1 day before the tumour inoculation, and repeated every 3 days. The following antibodies were used: anti-mouse CD4 (GK1.5, 100 μg); anti-mouse CD8a (YTS 169.4, 100 μg); anti-mouse Ly6G (1A8, 200 μg), all of which were obtained from BioXCell. For anti-CD80 treatment, 200 μg anti-mouse CD80 antibody (16-10A1, BioXCell) or isotype control (BE0091, BioXCell) were given intraperitoneally 1 day before the tumour inoculation, and repeated every 3 days.

hMoDCs

Human peripheral blood mononuclear cells (PBMCs) were collected from healthy donors' buffy coat (provided by the University Hospitals of Geneva (HUG)) using Ficoll-Paque Plus (Cytiva). Monocytes were isolated using the Classical Monocyte Isolation Kit (human, Miltenyi Biotec) according to the manufacturer's instructions. Cells were centrifuged and resuspended in complete medium (RPMI, 10% heat-inactivated FCS, 2 mM l-glutamine, 1% penicillin–streptomycin, 50 μM β-mercaptoethanol) plus 500 U ml^{-1} human GM-CSF and 250 U ml^{-1} human IL-4 (both from Miltenyi Biotec) and cultured in 12-well cell culture plates at 37 °C with 5% CO_2. The medium was refreshed every other day. On day 6, non-adherent cells were collected for experiments.

Proliferation assays

Mouse spleen and lymph nodes were collected from OT-I mice and then passed through a 70 μm cell strainer. After obtaining single-cell suspension, naive $CD8^+$ T cells were purified using a cell isolation kit (Miltenyi) according to the manufacturer's instructions. Cells were adjusted to 2×10^6 cells per ml concentration and stained using the Cell Trace Violet Proliferation Kit (C34557, Invitrogen) at a final concentration of 5 μM in PBS for 15 min at 37 °C. LPS-matured BMDCs were loaded with 10 nM OVA-peptide SIINFEKL for 15 min at 37 °C. Then, 1,000 BMDCs were co-cultured with 10,000 naive $CD8^+$ T cells in 200 μl complete medium (RPMI, 10% heat-inactivated FCS, 2 mM l-glutamine, 1% penicillin–streptomycin, 50 μM β-mercaptoethanol, 1 mM pyruvate sodium) in a 96-well round-bottom plate. Plates were incubated at 37 °C with 5% CO_2 for 48 h before analysis. For polyclonal proliferation assays, naive $CD8^+$ T cells were collected from C57BL/6N mice and labelled as described above. A total of 1,000 LPS-matured BMDCs were co-cultured with 10,000 naive $CD8^+$ T cells in 200 μl complete medium (RPMI, 10% heat-inactivated FCS, 2 mM l-glutamine, 1% penicillin–streptomycin, 50 μM β-mercaptoethanol, 1 mM pyruvate sodium) supplemented with an anti-mouse CD3 antibody (1 μg ml^{-1}, Thermo Fisher Scientific, 16-0032-82) to provide the TCR signal. hMDCs were synchronized using horse serum as described above. Before co-culture with T cells, hMDCs were matured with LPS (200 ng ml^{-1}) for 24 h. To assess proliferation, HLA-$A2^+$ hMoDCs were used and loaded with

Melan-$A_{26–35(A27L)}$ peptide (ELAGIGILTV (ELA), 10 μg ml^{-1})) 24 h before co-culture. For naive T cell proliferation assays, human $CD8^+$ T cells were isolated from PBMCs (the same donor as the hMoDCs) using a $CD8^+$ T cell isolation kit (Miltenyi Biotec) and then labelled with Cell Trace Violet (Invitrogen). A total of 10,000 labelled T cells were co-cultured with 1,000 matured hMoDCs, together with 5 μg ml^{-1} an anti-human CD3 antibody (Invitrogen, 16-0037-81). Then, 5 days later, T cells were collected for flow cytometry analysis. For patients' T cell proliferation assays, antigen-specific $CD8^+$ T cells were FACS-sorted from PBMCs of patients with melanoma using PE-conjugated HLA-A2/ELA multimers. $Multimer^+$ cells were cloned by limiting dilution and expanded with phytohaemagglutinin and allogenic feeder cells in a medium containing 150 U ml^{-1} human recombinant IL-2 (hrIL-2), as previously described[20]. Then, single clones of T cells were used for co-culture with hMoDCs at a ratio of 5:1. On day 5, cells were collected for flow cytometry analysis.

Vaccinations

Unless specified, 30 μg OVA together with 20 μg CpG OND 1826 and 20 μg poly(I:C) (VacciGrade, all from InvivoGen) were injected subcutaneously into tumour-bearing mice adjacent to the tumour. For vaccination with BMDCs, 1 million SIINFEKL-loaded LPS-matured BMDCs were injected together with 20 μg CpG and 20 μg poly(I:C). The tumour volume was then measured every one or two days using callipers. The human vaccination trial was performed as previously described[20]. The times of vaccination were stratified into vaccinations performed before or after 13:00. The patients included received all of their vaccinations before or after this cut-off time.

In vitro cell treatment

SIINFEKL-loaded, LPS-matured BMDCs were adjusted to a concentration of 1×10^6 cells per well and incubated with anti-mouse CD80 (50 μg ml^{-1}, 16-10A1, BioXCell) or isotype control for 15 min at 37 °C. Proliferation assays were then performed as described above.

Immunofluorescence imaging

B16-F10-OVA cells were cultured as described above, and labelled with CellTrace Violet (Invitrogen). Cells were counted and resuspended into PBS; an equal volume of cold Corning Matrigel was then added and mixed thoroughly with tumour cells. One million cells in 50 µl were injected subcutaneously into the right flank of the mouse. Then, 4 h later, Matrigels were collected directly into 4% PFA and stored at 4 °C for 4 h. Matrigels were kept in 30% sucrose (Sigma-Aldrich) overnight at 4 °C after fixation, embedded into OCT blocks (CellPath) and kept at −80 °C. Matrigels were subsequently dissected and processed for cryosectioning with 50 µm serial cryosections being cut and processed for immunohistochemistry. The sections were post-fixed with 4% PFA for 10 min at room temperature. After three washes with PBS, the sections were incubated with blocking buffer (PBS with 20% normal goat serum and 0.5% Triton X-100) for 2 h at room temperature. After three consecutive washes with PBS, the sections were stained with an antibody mix of FITC conjugated mouse anti-CD45 (30-F11; BioLegend), PE/Dazzle594-conjugated mouse anti-CD11b (M1/70; BioLegend), Alexa Fluor F647-conjugated mouse anti-CD11c (N418, BioLegend) diluted in the same blocking buffer as before and incubated overnight at 4 °C. The sections were washed three times with PBS before mounting in Fluoromount Aqueous Mounting Medium (Sigma-Aldrich). Images of Matrigels were obtained as sections using the Zeiss Axio Examiner.Z1 confocal spinning-disk microscope equipped with 405-, 488-, 561- and 640-nm laser sources. Step size was determined as 4 µm and images were acquired at ×20 magnification. All image analyses were performed in ImageJ. Volume fractions were obtained from binary images in a 3D environment by thresholding the voxels for both Matrigel volume and the signals of interest.

Sorting of $CD11c^{+}MHCII^{+}$ cells and RNA-seq

To obtain DCs, draining inguinal lymph nodes were collected from mice 24 h after tumour engraftment and collected at four timepoints (ZT3, ZT9, ZT15 and ZT21; $n = 5$ mice for control, $n = 3$ mice for *Clec9acre:Bmal1*^{flox} and $n = 3$ mice for sham-injected mice). Lymph nodes were digested as previously described, and $CD45^{+}CD11c^{+}MHCII^{high}$ cells were sorted using an Astrios sorter (Beckman). Flow-cytometry-sorted DCs were collected in RNAprotect Cell Reagent (76526, Qiagen). RNA was isolated using the

RNeasy Plus Micro Kit (74034, Qiagen) according to the manufacturer's instructions. RNA integrity and quantity were assessed using a Bioanalyzer (Agilent Technologies). cDNA libraries were constructed by the Genomic platform of the University of Geneva as follows: the SMART-Seq v.4 kit from Clontech was used for the reverse transcription and cDNA amplification according to the manufacturer's specifications, starting with 1 ng of total RNA as the input. Then, 200 pg of cDNA was used for library preparation using the Nextera XT kit from Illumina. Library molarity and quality were assessed using the Qubit and Tapestation using a DNA High sensitivity chip (Agilent Technologies). Libraries were pooled and loaded for clustering on two lanes of a single-read Illumina flow cell. Reads of 50 bases were generated using the TruSeq SBS chemistry on the Illumina HiSeq 4000 sequencer.

Reads were aligned using STAR (v.2.7.0)[22] to the mouse mm10 UCSC genome. Gene expression was quantified using HTSeq (v.0.9.1)[23]. Differential expression analysis was performed using the R/Bioconductor edgeR package. The counts were normalized according to the library size and filtered. Genes with counts above 1 count per million reads in at least five samples were retained for subsequent analysis. Tests for differentially expressed genes were performed using a general linear model using a negative binomial distribution. The genes were considered to be differentially expressed when the fold change was at least twofold with a 5% false-discovery rate (FDR) Benjamini–Hochberg multiple-testing correction. DiscoRhythm[24] R package (v.1.10.0) was used to characterize the rhythmicity present in the provided dataset by performing outlier detection, principal component analysis (PCA) and detection of gene-wise oscillation characteristics. The default parameters were used, except when indicated.

PCA was used to extract the strongest recurring patterns in the dataset. Gene expression values were scaled to a s.d. of one before PCA, such that all genes were on an equal scale. The first four PCA scores were used to detect outliers (flagged by their deviation from the mean). A threshold of three units of s.d. was used. The cosinor method was used to test the summarized temporal signal for rhythmicity. PC1 and PC2 were retained as they scored above 10% of the variance each (WT: PC1, 18.9%; PC2, 13.6%; *Bmal1*$^{\Delta cDC}$: PC1, 20.6%; PC2, 12.3%; PBS sham injection: PC1, 20.2%; PC2, 11.6%)

suggesting that two main phases of oscillation exist in the data (Supplementary Tables 1–3). Each gene was tested for rhythmicity with a significance value of $P < 0.05$. Genes with significant rhythmicity were assigned to two sets depending on their acrophase (the time in a periodic cycle during which a temporal pattern is at its maximum value). Genes with a sincoef > 0 corresponded to an acrophase between 0 and 12 h (PC1), whereas genes with a sincoef < 0 corresponded to an acrophase between 12 and 24 h (PC2). Both oscillating gene lists were tested for pathway enrichment, using over-representation analyses in the Gene Ontology Biological Process (GOBP) and Reactome pathways, using ClusterProfiler R[25,26] package (v.4.4.4). Pathways with an enrichment $P < 0.05$ were reported as significant.

ChIP with qPCR

A total of 2×10^7 BMDCs were collected and fixed in PBS containing 1% formaldehyde (Thermo Fisher Scientific) for 10 min at room temperature and quenched with 1 M glycine in PBS. Cells were then pelleted and sonicated (Diagenode Bioruptor) to obtain fragments of 0.2–0.8 kilobases in size. Immunoprecipitation was performed using anti-BMAL1 (D2L7G, Cell Signaling Technology), anti-histone H3 (Abcam) or control IgG (Cell Signaling Technology). DNA was isolated using the MinElute PCR Purification kits (Qiagen). qPCR was performed using PowerUp SYBR Green (Applied Biosystems) in the StepOne Real-Time PCR System. Occupancy of BMAL1 at the *Cd80* and *Per2* promoters was quantified by qPCR targeting regions identified as containing E-boxes using the SCOPE motif finder and EPFL eukaryotic database. Relative enrichment was determined as the percentage of input.

Statistical analyses

Unless otherwise specified, all data were plotted from independent biological replicates. Data were analysed using Prism 9 (GraphPad). $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$. Unless otherwise specified, Student's *t*-tests were two-tailed. All other statistical information, including *t*, *F* and d.o.f. values, are provided in the source data.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data supporting the conclusions of this paper are available online (https://doi.org/10.26037/yareta:t47xfgrgyvbi3kkg7lfidvzw2q). The sequencing data have been deposited at the NCBI Gene Expression Omnibus (GEO) and are accessible under GEO series accession number GSE217381. Source data are provided with this paper.

References

1. Gajewski, T. F., Schreiber, H. & Fu, Y. X. Innate and adaptive immune cells in the tumor microenvironment. *Nat. Immunol.* **14**, 1014–1022 (2013).

2. Chen, D. S. & Mellman, I. Elements of cancer immunity and the cancer-immune set point. *Nature* **541**, 321–330 (2017).

3. Arjona, A., Silver, A. C., Walker, W. E. & Fikrig, E. Immunity's fourth dimension: approaching the circadian-immune connection. *Trends Immunol* **33**, 607–612 (2012).

4. Curtis, A. M., Bellet, M. M., Sassone-Corsi, P. & O'Neill, L. A. Circadian clock proteins and immunity. *Immunity* **40**, 178–186 (2014).

5. Labrecque, N. & Cermakian, N. Circadian Clocks in the Immune System. *J. Biol. Rhythms* **30**, 277–290 (2015).

6. Man, K., Loudon, A. & Chawla, A. Immunity around the clock. *Science* **354**, 999–1003 (2016).

7. Pick, R., He, W., Chen, C. S. & Scheiermann, C. Time-of-day-dependent trafficking and function of leukocyte subsets. *Trends*

Immunol. **40**, 524–537 (2019).

8. Scheiermann, C., Gibbs, J., Ince, L. & Loudon, A. Clocking in to immunity. *Nat. Rev. Immunol.* **18**, 423–437 (2018).

9. Scheiermann, C., Kunisaki, Y. & Frenette, P. S. Circadian control of the immune system. *Nat. Rev. Immunol.* **13**, 190–198 (2013).

10. Palomino-Segura, M. & Hidalgo, A. Circadian immune circuits. *J. Exp. Med.* **218**, e20200798 (2021).

11. Druzd, D. et al. Lymphocyte circadian clocks control lymph node trafficking and adaptive immune responses. *Immunity* **46**, 120–132 (2017).

12. Fortier, E. E. et al. Circadian variation of the response of T cells to antigen. *J Immunol* **187**, 6291–6300 (2011).

13. Silver, A. C., Arjona, A., Walker, W. E. & Fikrig, E. The circadian clock controls toll-like receptor 9-mediated innate and adaptive immunity. *Immunity* **36**, 251–261 (2012).

14. Sutton, C. E. et al. Loss of the molecular clock in myeloid cells exacerbates T cell-mediated CNS autoimmune disease. *Nat. Commun.* **8**, 1923 (2017).

15. Long, J. E. et al. Morning vaccination enhances antibody response over afternoon vaccination: a cluster-randomised trial. *Vaccine* **34**, 2679–2685 (2016).

16. de Bree, L. C. J. et al. Circadian rhythm influences induction of trained immunity by BCG vaccination. *J. Clin. Invest.* **130**, 5603–5617 (2020).

17. Papagiannakopoulos, T. et al. Circadian rhythm disruption promotes lung tumorigenesis. *Cell Metab.* **24**, 324–331 (2016).

18. Balsalobre, A., Damiola, F. & Schibler, U. A serum shock induces circadian gene expression in mammalian tissue culture cells. *Cell* **93**, 929–937 (1998).

19. Holtkamp, S. J. et al. Circadian clocks guide dendritic cells into skin lymphatics. *Nat. Immunol.* **22**, 1375–1381 (2021).

20. Speiser, D. E. et al. Rapid and strong human CD8+ T cell responses to vaccination with peptide, IFA, and CpG oligodeoxynucleotide 7909. *J. Clin. Invest.* **115**, 739–746 (2005).

21. Lutz, M. B. et al. An advanced culture method for generating large quantities of highly pure dendritic cells from mouse bone marrow. *J. Immunol. Methods* **223**, 77–92 (1999).

22. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

23. Anders, S., Pyl, P. T. & Huber, W. HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

24. Carlucci, M. et al. DiscoRhythm: an easy-to-use web application and R package for discovering rhythmicity. *Bioinformatics* **36**, 1952–1954 (2019).

25. Wu, T. et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021).

26. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).

Acknowledgements

We thank B. Schraml and C. Reis e Sousa for providing *Cle9a-cre* mice; S. Jemelin for help with genotyping; F. Sinturel for help with ChIP; the staff at the IGE3 Genomics Platform and the Flow Cytometry Core Facility of the University of Geneva; L. Cheng for help with transduction; and S. Xu for discussions. This study received support from the European Research Council (ERC CoG 101001233, CIRCADYN, to C.S.), the Swiss National Science Foundation (SNSF, 310030_182417/1, to C.S., and 310030_185255,

to S. Hugues), the Swiss Cancer League (KLS-4836-08-2019, to C.S.), the Geneva Cancer League (2106, to C.S.), the EU ITN (813284, INTEGRATA, to C.S.) and the Novartis Foundation for medical-biological Research (20A019, to C.S.). C.W. is a recipient of a fellowship of the Chinese Scholarship Council (CSC). A.N. and I.C. acknowledge the support of the Italian Association for Cancer Research (IG#22098, to A.N., and MFAG#26482, to I.C.). C.J. is supported by a SNSF PRIMA fellowship (PR00P3_179727), the Swiss Cancer League (KFS 5250-02-2021) and the Geneva Cancer League (GCL, 2007). M.C is the recipient of an iGE3 PhD salary award. M.J.P. receives support from the ISREC Foundation, Ludwig Cancer Research, and NIH grants P01-CA240239 and R01-CA218579. R.B. was funded by a Postdoc Mobility fellowship and a Return Grant of the SNSF (P400PM_183852, P5R5PM_203164).

Author information

Authors and Affiliations

1. Department of Pathology and Immunology, Faculty of Medicine, University of Geneva, Geneva, Switzerland

   Chen Wang, Coline Barnoud, Mara Cenerenti, Mengzhu Sun, Burak Kizil, Ruben Bill, Robert Pick, Laure Garnier, Olga A. Gkountidi, Louise M. Ince, Stéphanie Hugues, Mikaël J. Pittet, Camilla Jandus & Christoph Scheiermann

2. Ludwig Institute for Cancer Research, Lausanne, Switzerland

   Mara Cenerenti, Mikaël J. Pittet & Camilla Jandus

3. Department of Internal Medicine and Medical Specialties, University of Genoa, Genoa, Italy

   Irene Caffa & Alessio Nencioni

4. AGORA Cancer Research Center, Lausanne, Switzerland

   Ruben Bill, Olivier Michielin & Mikaël J. Pittet

5. Center for Systems Biology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

Ruben Bill & Mikaël J. Pittet

6. Department of Computational Biology, University of Lausanne, Lausanne, Switzerland

Yuanlong Liu

7. Swiss Cancer Center Leman, Lausanne, Switzerland

Yuanlong Liu

8. Translational Data Science (TDS), Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland

Yuanlong Liu & Nadine Fournier

9. Biomedical Center (BMC), Institute for Cardiovascular Physiology and Pathophysiology, Walter-Brendel-Center for Experimental Medicine (WBex), Faculty of Medicine, Ludwig-Maximilians-Universität Munich, Planegg-Martinsried, Germany

Stephan Holtkamp & Christoph Scheiermann

10. Department of Oncology, University of Lausanne, Lausanne, Switzerland

Daniel E. Speiser

11. Geneva Centre for Inflammation Research, Geneva, Switzerland

Stéphanie Hugues, Camilla Jandus & Christoph Scheiermann

12. IRCCS Ospedale Policlinico San Martino, Genoa, Italy

Alessio Nencioni

Contributions

Conceptualization: C.S., C.W., C.J., M.J.P., A.N., S. Hugues, D.E.S. and O.M. Methodology: C.W., B.K., I.C., R.B., L.M.I. and N.F. Investigation: C.W., C.B., B.K., M.C., I.C., R.B., M.S., L.G., O.A.G., L.M.I. and S. Holtkamp. Visualization: Y.L., R.P. and N.F. Supervision: C.S. Writing: C.W., C.B., B.K., M.J.P. and C.S.

Corresponding author

Correspondence to Christoph Scheiermann.

Ethics declarations

Competing interests

M.J.P. has been a consultant for AstraZeneca, Debiopharm, Elstar Therapeutics, ImmuneOncia, KSQ Therapeutics, MaxiVax, Merck, Molecular Partners, Third Rock Ventures and Tidal; these relationships are unrelated to the current study. The other authors declare no competing interests.

Peer review

Peer review information

Nature thanks Achim Kramer, Thales Papagiannakopoulos and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Time of day of engraftment dictates tumour size.

(**a**) Tumour volume of each mouse after engraftment of B16-F10-OVA cells at 6 different times of the day (*Zeitgeber* time (ZT)); n = 10 mice per timepoint from 2 independent experiments. (**b**) Tumour volume after engraftment of E0771 cells at 2 different times of the day, plotted per group (left) and mouse (right); n = 5 mice from 2 independent experiments, two-way ANOVA. (**c-d**) Normalized tumour volume, plotted per group (left) and mouse (right) (**c**) and tumour volume on day 20 (**d**), after engraftment of 4T1 cells at 2 different times of the day; n = 12 (ZT9), n = 10 (ZT21) mice from 2 independent experiments, two-way ANOVA (**c**) and unpaired Student's *t*-test (**d**). (**e**) Normalized tumour volume after engraftment of MC-38 cells at 2 different times of the day, plotted per group (left) and mouse (right); n = 8 (ZT9), n = 6 (ZT21) mice from 2 independent experiments, two-way ANOVA. (**f**) Fluorescence flux in photons/seconds of B16-F10-OVA-Luc (5×10^5 cells) tumours on day 5 after engraftment; n = 8 (ZT9), n = 7 (ZT21) mice from 2 independent experiments, unpaired Student's *t*-test. (**g**) Tumour volume in phase-shifted mice injected using the same batch of B16-F10 cells (without OVA expression); n = 9 (ZT9), n = 6 (ZT21) mice from 2 independent experiments, two-way ANOVA. (**h**) Scheme of the normal (light:dark), inverted lighting (dark:light) and jet lag protocols. For the jet lag, every three days mice were placed into a 6 h- or 12 h-phase delayed environment. The red dots represent the time when tumours were engrafted (9 h after the start of the cycle). All data are represented as mean ± SEM, all *t*-tests are two-tailed.

Source data

Extended Data Fig. 2 Depletion of CD4 T cells and neutrophils.

(**a**) Flow cytometry gating strategy of tumour infiltrating leukocytes. (**b**) Normalized cell numbers of tumour infiltrating leukocytes after 14 days of tumour engraftment; from ZT1 to 21, n = 10, 9, 10, 7, 10, 8 mice from 4 independent experiments, Cosinor analysis. Treg, regulatory T cells. (**c-d**)

Tumour volume upon anti-CD4 antibody depletion, n = 4 (ZT9 anti-CD4), n = 5 (ZT21 anti-CD4), n = 6 (control) mice (**c**), or anti Ly6G antibody depletion, n = 5 (control), n = 6 (anti-Ly6G) mice (**d**), from 2 independent experiments, two-way ANOVA. (**e**-**j**) Flow cytometry gating strategies and quantification of neutrophils in blood (**e**-**h**) or tumour (**i**-**j**) after anti-Ly6G treatment. Anti-mouse Ly6G antibody was given every 3 days, starting 1 day before the tumour inoculation (d-1). Neutrophil frequency after treatment at days 0 and 1 (**e**-**f**), from left to right, n = 3, 2, 3 mice, or day 12 (**g**-**j**), n = 3 (iso), n = 5 (anti-Ly6G) mice from 2 independent experiments, unpaired Student's *t*-test. All data are represented as mean ± SEM, ns, not significant, all *t*-tests are two-tailed.

Source data

Extended Data Fig. 3 Leukocyte populations in skin and dLN.

(**a**) Flow cytometry gating strategy of skin myeloid populations. (**b**-**c**) Number (**b**) and proportion (**c**) of leukocytes at the tumour engraftment site 4 h after B16-F10-OVA engraftment; n = 8 mice from 2 independent experiments, unpaired Student's *t*-test. Leu, leukocyte; N, neutrophils; EOS, eosinophil; IM, inflammatory monocytes; NK, natural killer cells. (**d**) Flow cytometry gating strategy of DC subsets in draining lymph nodes (dLN). (**e**-**i**) Leukocytes, T cell (**e**-**f**) and DCs (**g**-**h**) in the dLN, 24 h after tumour engraftment (5×10^5 B16-F10-OVA); n = 8 mice, from 2 independent experiments, unpaired Student's *t*-test. (**i**) Gating of anti-mouse H-2K^b bound to SIINFEKL$^+$ DCs in the B16-F10-OVA model. All data are represented as mean ± SEM, all *t*-tests are two-tailed.

Source data

Extended Data Fig. 4 Time-of-day differences exist in other tumour models.

(**a**-**d**) DCs and T cell numbers in the dLN, 24h after tumour engraftment of MC-38 cells (n = 6 mice) (**a**), 4T1 cells (n = 6 mice) (**b**), or E0771-OVA cells, n = 7 (ZT9), n = 6 (ZT21) mice, (**c**-**d**), from 2 independent experiments, unpaired Student's *t*-test. (**c**) Gating of anti-mouse H-2K^b

bound to SIINFEKL^{+} DCs in the E0771-OVA model. (**e**-**g**) T cell numbers in the dLN 72h after tumour engraftment of E0771-OVA cells, n = 5 (ZT9), n = 7 (ZT21) (**e**), or MC-38 cells, n = 6 mice (**f**-**g**), from 2 independent experiments, unpaired Student's *t*-test. (**f**) Gating of Dextramer H-2D^{b}/ ASMTNMELM bound to CD8 T cells in the MC-38 model and, (**g**) quantification. All data are represented as mean ± SEM, all *t*-tests are two-tailed.

Source data

Extended Data Fig. 5 Differences in DCs and CD4 T cells in the dLN.

(**a**) DCs numbers in the dLN 24h after sham PBS injection without tumour inoculation, n = 7 (ZT9), n = 8 (ZT21) mice, from 2 independent experiments, unpaired Student's *t*-test. (**b**) DQ-OVA^{+} DCs in dLN 24h after inoculation; n = 4 mice from 2 independent experiments, unpaired Student's *t*-test. (**c**) EdU staining in CD3^{+}CD4^{+} T cells in the dLN 48 h after B16-F10-OVA cell engraftment; n = 3 (ZT9), n = 4 (ZT21) mice, representative from 2 independent experiments, unpaired Student's *t*-test. All data are represented as mean ± SEM, all *t*-tests are two-tailed.

Source data

Extended Data Fig. 6 RNA-seq analyses of CD11^{+} MHCII^{hi} cells in the dLN.

(**a**-**d**) RNA-seq analyses of CD11c^{+} MHCII^{hi} cells in the dLN 24h after B16-F10-OVA cell engraftment in control mice (n = 5 mice) or *Clec9acre:Bmal1^{flox}* mice (n = 3 mice), from 2 independent experiments. (**a**) Expression (Counts per million (CPM)) of *Per1* and *Dbp* in CD11c^{+} MHCII^{hi} cells from control mice, Cosinor analysis. (**b**) Principal Component (PC) analyses of each sample from sequencing of CD11c^{+} MHCII^{hi} DCs in control mice. (**c**) Significantly enriched pathways from PC1 in control cells with CD28 signalling pathways highlighted in red, shown for control and *Clec9acre:Bmal1^{flox}* CD11c^{+} MHCII^{hi} cells. The vertical dashed line

represents the significant *p* values, hypergeometric test. (**d**) GO Biological Process interactions in the PC2 gene cluster for control cells. All data are represented as mean ± SEM.

Source data

Extended Data Fig. 7 RNA-seq analyses in *Clec9acre:Bmal1^{flox}* mice.

(**a**-**c**) RNA-seq analyses of CD11c^{+} MHCII^{hi} cells in the dLN 24h after B16-F10-OVA cell engraftment at ZT3, 9, 15 and ZT21 in *Clec9acre:Bmal1^{flox}* mice (n = 3 mice). (**a**-**b**) Significantly enriched pathways from PC1 (**a**) and PC2 (**b**) in *Clec9acre:Bmal1^{flox}* cells, hypergeometric test. (**c**) GO Biological Process interactions in PC2 gene cluster for *Clec9acre:Bmal1^{flox}* cells.

Extended Data Fig. 8 RNA-seq analyses in sham conditions.

(**a**-**c**) RNA-seq analyses of CD11c^{+} MHCII^{hi} cells in the dLN 24h after PBS injection (n = 3 mice) or after B16-F10-OVA cell (n = 5 mice) at ZT3, 9, 15 and ZT21 in WT mice, from 2 independent experiments. (**a**) Pathways found significantly enriched by over-representation analysis in the lists of significantly oscillating genes in PC1 (Reactome database) in WT mice. The same pathways from RNA-seq analyses in PBS injection mice were also plotted; n = 3 mice per timepoint. The vertical dashed line represents the significant *P* values, hypergeometric test. (**b**-**c**) Pathways found significantly enriched by over-representation analysis in the lists of significantly oscillating genes in PC1 and PC2 in PBS injection mice, hypergeometric test.

Extended Data Fig. 9 Synchronization experiments of BMDCs.

(**a**-**b**) Synchronization scheme of BMDCs for qPCR analyses (**a**) and co-culture experiments (**b**). The schematics in **a** and **b** were created using BioRender. (**c**) *Cd80* mRNA expression after synchronization of immature BMDCs from WT (n = 15 mice) and *Bmal1$^{-/-}$* (n = 4 mice) mice without LPS maturation, Cosinor analysis. (**d**) qPCR of LPS-matured BMDCs at

different times after synchronization; n = 9 mice from 2 independent experiments, one-way ANOVA. (**e**) Predicted binding regions of BMAL1 to the *Cd80* gene using Eukaryotic Promoter Database with a cutoff *P*-value of 0.001. (**f**) Chromatin immunoprecipitation (ChIP) of BMAL1 binding the promoter of *Per2* of BMDCs after synchronization; n = 3 mice, from 2 independent experiments, two-way ANOVA. All data are represented as mean ± SEM, ns, not significant.

Source data

Extended Data Fig. 10 Time-of-day differences in vaccination efficacy.

(**a**) Tumour volume after B16-F10-OVA cell engraftment at ZT9 and OVA vaccination at ZT9 (120 h after tumour engraftment, n = 11) or ZT21 (108 h, n = 12, or 132 h, n = 11 mice, after tumour engraftment); from 2 independent experiments, two-way ANOVA. (**b**-**c**) Numbers of DC subsets (**b**) and T cells (**c**) in the draining LN 24h after OVA vaccination (on day 5 after B16-F10-OVA cell engraftment) in control or *Clec9acre:Bmal1*^{flox} mice, n = 5 mice from 2 independent experiments, unpaired Student's *t*-test. (**d**) Tumour volume in WT mice after tumour engraftment (5×10^5 B16-F10-OVA cells) at ZT9 or ZT21, with or without OVA vaccination on day 5 (arrow) at ZT9 or ZT21, n = 4 mice, two-way ANOVA. (**e**) Tumour volume after B16-F10-OVA cell engraftment at ZT9 and OVA vaccination on day 5 and 8 (arrows), both at ZT9 or ZT21 (n = 6 mice) or unvaccinated controls (n = 9 mice), from 2 independent experiments, two-way ANOVA. (**f**) qPCR of human *PER2* expression in human monocyte derived DCs (hMoDC) after synchronization, n = 3 patients, Cosinor analysis. All data are represented as mean ± SEM, ns, not significant, all *t*-tests are two-tailed.

Source data

Supplementary information

Supplementary Figures

Supplementary Figs. 1–3, showing gating strategies relevant to the Figures and Extended Data Figures.

Reporting Summary

Supplementary Table 1

Full lists of genes significantly ($P < 0.05$) oscillating in CD11c^{+}MHCII^{high} cells collected from WT mice in the dLN collected 24 h after subcutaneous B16F10-OVA tumour engraftment in the skin at ZT9 or ZT21 in the PC1 compartment (acrophase < 12), and the PC2 compartment (acrophase > 12). Reactome pathways significantly enriched in PC1 genes, GO biological process pathways in PC2 genes ($P < 0.05$).

Supplementary Table 2

Full lists of genes significantly ($P < 0.05$) oscillating in CD11c^{+}MHCII^{high} cells collected from *Bmal1*$^{\Delta cDC}$ mice in the dLN collected 24 h after subcutaneous B16F10-OVA tumour engraftment in the skin at ZT9 or ZT21 in the PC1 compartment (acrophase < 12), and the PC2 compartment (acrophase > 12). Reactome pathways significantly enriched in PC1 genes, GO biological process pathways in PC2 genes ($P < 0.05$).

Supplementary Table 3

Full lists of genes significantly ($P < 0.05$) oscillating in CD11c^{+}MHCII^{high} cells collected from WT mice in the dLN collected 24 h after subcutaneous sham injection (PBS) in the skin at ZT9 or ZT21 in the PC1 compartment (acrophase < 12), and the PC2 compartment (acrophase > 12). Reactome pathways significantly enriched in PC1 genes, GO biological process pathways in PC2 genes ($P < 0.05$).

Supplementary Table 4

Primers used in this study.

Source data

Source Data Fig. 1

Source Data Fig. 2

Source Data Fig. 4

Source Data Fig. 5

Source Data Extended Data Fig. 1

Source Data Extended Data Fig. 2

Source Data Extended Data Fig. 3

Source Data Extended Data Fig. 4

Source Data Extended Data Fig. 5

Source Data Extended Data Fig. 6

Source Data Extended Data Fig. 9

Source Data Extended Data Fig. 10

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the

material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Wang, C., Barnoud, C., Cenerenti, M. *et al.* Dendritic cells direct circadian anti-tumour immune responses. *Nature* **614**, 136–143 (2023). https://doi.org/10.1038/s41586-022-05605-0

- Received: 14 April 2022
- Accepted: 25 November 2022
- Published: 05 December 2022
- Issue Date: 02 February 2023
- DOI: https://doi.org/10.1038/s41586-022-05605-0

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article is cited by

- **Dendritic cells rise and shine**
    - Joseph Willson

  Nature Reviews Cancer (2023)

- **Time of day shapes the success of a cancer treatment**
    - Christian H. Gabriel
    - Achim Kramer

  Nature (2023)

Dendritic cells rise and shine

- Joseph Willson

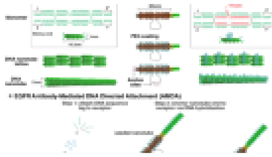
Nature Reviews Cancer Research Highlight 21 Dec 2022

Time of day shapes the success of a cancer treatment

- Christian H. Gabriel
- Achim Kramer

Nature News & Views 17 Jan 2023

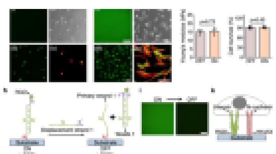
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05605-0



Growth and site-specific organization of micron-scale biomolecular devices on living mammalian cells

30 September 2021

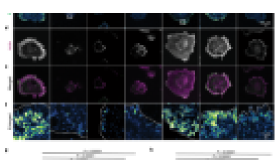
Sisi Jia, Siew Cheng Phua, … Rebecca Schulman



Programmable integrin and N-cadherin adhesive interactions modulate mechanosensing of mesenchymal stem cells by cofilin phosphorylation

11 November 2022

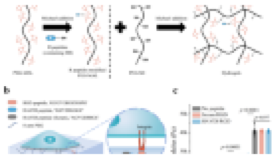
Zheng Zhang, Baoyong Sha, … Min Lin



Integrin nanoclusters can bridge thin matrix fibres to form cell–matrix adhesions

02 September 2019

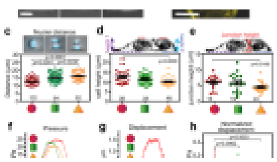
Rishita Changede, Haogang Cai, … Michael P. Sheetz



[Mechanics-driven nuclear localization of YAP can be reversed by N-cadherin ligation in mesenchymal stem cells]

28 October 2021

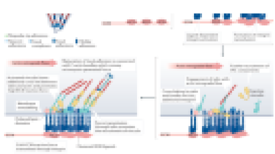
Cheng Zhang, Hongyuan Zhu, … Min Lin



[Intrinsic cell rheology drives junction maturation]

17 August 2022

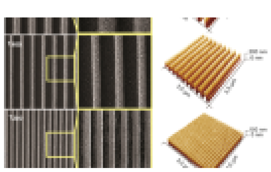
K. Sri-Ranjan, J. L. Sanchez-Alonso, … Vania M. M. Braga



[Organization, dynamics and mechanoregulation of integrin-mediated cell–ECM adhesions]

27 September 2022

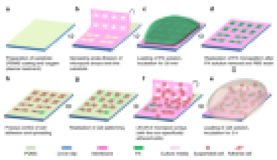
Pakorn Kanchanawong & David A. Calderwood



[Neuronal contact guidance and YAP signaling on ultra-small nanogratings]

28 February 2020

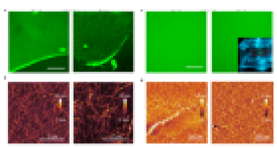
Ilaria Tonazzini, Cecilia Masciullo, … Marco Cecchini



[A micropore array-based solid lift-off method for highly efficient and controllable cell alignment and spreading]

07 September 2020

Tingting Hun, Yaoping Liu, … Wei Wang



[Adaptive liquid interfaces induce neuronal differentiation of mesenchymal stem cells through lipid raft assembly]

03 June 2022

Xiaofang Jia, Jingwen Song, … Katsuhiko Ariga

- Article
- Open Access
- Published: 12 December 2022

Programming multicellular assembly with synthetic cell adhesion molecules

- Adam J. Stevens,
- Andrew R. Harris,
- Josiah Gerdts,
- Ki H. Kim,
- Coralie Trentesaux,
- Jonathan T. Ramirez,
- Wesley L. McKeithan,
- Faranak Fattahi,
- Ophir D. Klein,
- Daniel A. Fletcher &
- …
- Wendell A. Lim

Show authors

Nature **volume 614**, pages 144–152 (2023)

- 15k Accesses
- 1 Citations
- 444 Altmetric
- Metrics details

Abstract

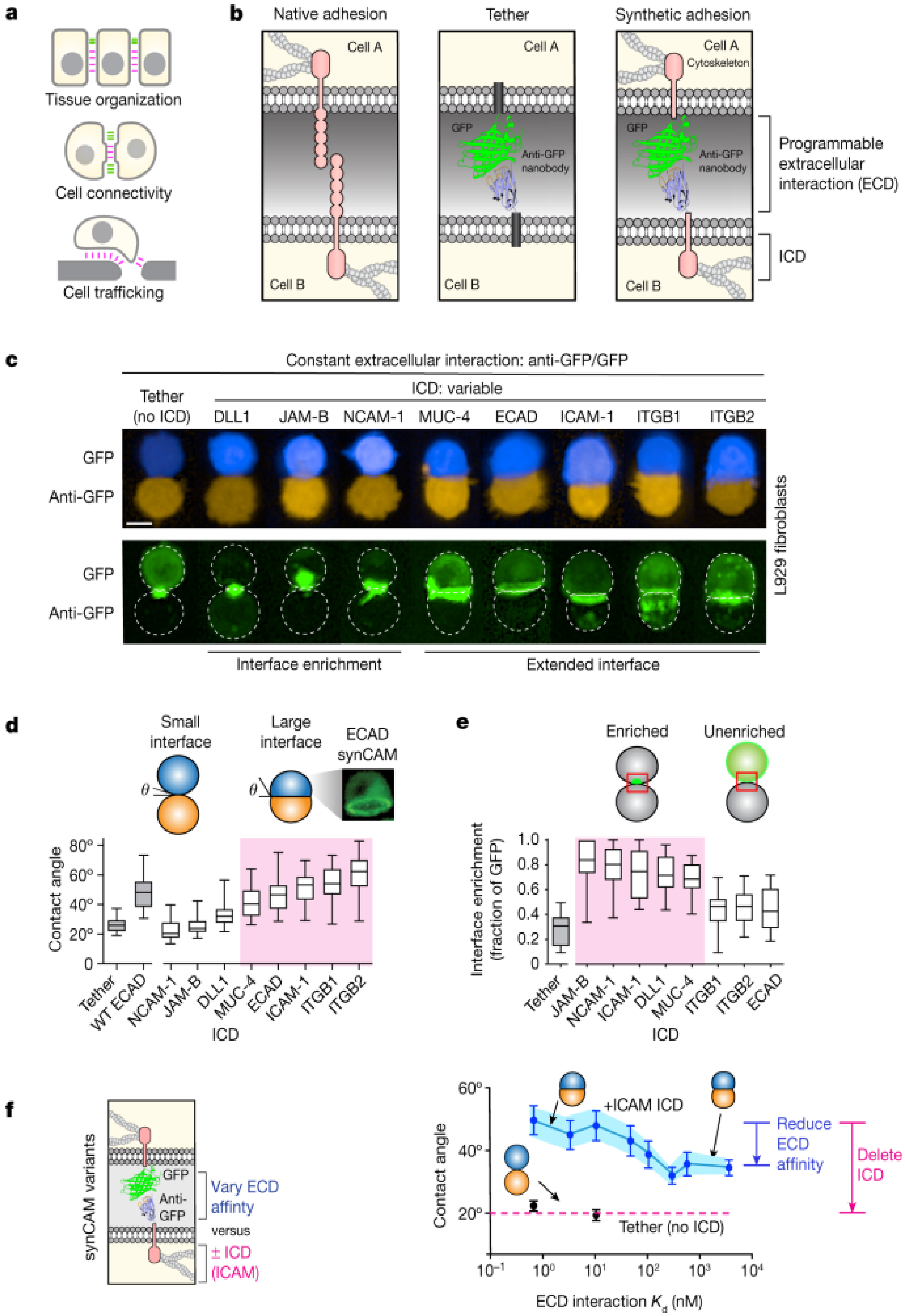
Cell adhesion molecules are ubiquitous in multicellular organisms, specifying precise cell–cell interactions in processes as diverse as tissue development, immune cell trafficking and the wiring of the nervous

system[1,2,3,4]. Here we show that a wide array of synthetic cell adhesion molecules can be generated by combining orthogonal extracellular interactions with intracellular domains from native adhesion molecules, such as cadherins and integrins. The resulting molecules yield customized cell–cell interactions with adhesion properties that are similar to native interactions. The identity of the intracellular domain of the synthetic cell adhesion molecules specifies interface morphology and mechanics, whereas diverse homotypic or heterotypic extracellular interaction domains independently specify the connectivity between cells. This toolkit of orthogonal adhesion molecules enables the rationally programmed assembly of multicellular architectures, as well as systematic remodelling of native tissues. The modularity of synthetic cell adhesion molecules provides fundamental insights into how distinct classes of cell–cell interfaces may have evolved. Overall, these tools offer powerful abilities for cell and tissue engineering and for systematically studying multicellular organization.

Main

The ability to systematically program cell–cell adhesion would provide powerful new tools to study development, neurobiology and immunology, and could facilitate the repair of multicellular tissues and the design of therapeutic cells[5,6] (Fig. 1a). Nonetheless, engineering adhesion in metazoan cells remains an underexplored area in synthetic biology.

Fig. 1: synCAMs facilitate custom cell–cell interactions.



a
Tissue organization
Cell connectivity
Cell trafficking
b
Native adhesion
Tether
Synthetic adhesion
Cell A
Cell B
Cytoskeleton
GFP
Anti-GFP nanobody
Programmable extracellular interaction (ECD)
ICD
c
Constant extracellular interaction: anti-GFP/GFP
ICD: variable
Tether (no ICD)
DLL1
JAM-B
NCAM-1
MUC-4
ECAD
ICAM-1
ITGB1
ITGB2
GFP
Anti-GFP
L929 fibroblasts
Interface enrichment
Extended interface
d
Small interface
Large interface
ECAD synCAM
Contact angle
ICD
e
Enriched
Unenriched
Interface enrichment (fraction of GFP)
ICD
f
synCAM variants
Vary ECD affinity
versus
± ICD (ICAM)
+ICAM ICD
Tether (no ICD)
Reduce ECD affinity
Delete ICD
ECD interaction K_d (nM)

a, Diverse functional roles of cell adhesion. **b**, The conceptual design of synCAM receptors. The extracellular domain of a CAM (left) is replaced by GFP and a GFP-binding nanobody (anti-GFP; right). A tether control lacking an ICD is also shown (middle). **c**, Maximum projection of ×20 confocal microscopy images of pairwise synCAM interfaces. Scale bar, 10 µm. $t = 3$ h. A GFP-expressing cell (blue) is bound to an anti-GFP-expressing cell (orange). The CAM TM and ICD domain for each pair is indicated (tether is the control lacking the ICD) (top). Bottom, the GFP channel of the interfaces above, highlighting the differences in receptor enrichment at the interface. Matched synCAM expression levels are shown in Extended Data Fig. 1. **d**, The contact angles measured from the interfaces shown in **a**. $n = 20$ (tether), $n = 20$ (WT ECAD), $n = 20$ (DLL1), $n = 20$ (JAM-B), $n = 20$ (NCAM-1), $n = 20$ (ICAM-1), $n = 20$ (ECAD), $n = 20$ (ITGB1), $n = 20$ (ITGB2), $n = 15$ (MUC4). Contact angles for WT ECAD homotypic cell–cell interaction are also shown. **e**, The fraction of GFP enrichment at the cell–cell interface from **c**. $n = 20$ (tether), $n = 20$ (DLL1), $n = 20$ (JAM-B), $n = 20$ (NCAM-1), $n = 20$ (ICAM-1), $n = 20$ (ECAD), $n = 20$ (ITGB1), $n = 20$ (ITGB2), $n = 15$ (MUC4). **f**, Quantification of the contact angles of pairwise L929 cells expressing GFP/anti-GFP synCAMs with the indicated affinities and in the presence (blue) or absence (black) of an ICAM-1 ICD. $n = 20$ pairs. The error bars show the 95% confidence intervals. $t = 3$ h. Matched synCAM expression levels are shown in Extended Data Fig. 1. An alternative analysis (competition cell sorting assay) of the same series of altered affinity synCAM cells is shown in Extended Data Fig. 3. For the box plots in **d** and **e**, the centre line shows the median, the box limits show the 25th to 75th percentile and the whiskers show the minimum to maximum values.

Source data

Native cell–cell interactions are mediated by a large collection of cell adhesion molecules (CAMs)—complex transmembrane proteins that bind neighbouring cells or matrix and induce a mechanical adhesive response, often involving cytoskeletal rearrangements[7,8,9,10,11]. Examples of CAMs include integrins, which assemble focal adhesions, and cadherins, which assemble adherens junctions between epithelial cells[11,12,13,14]. The structural complexity and functional diversity of CAMs make it unclear whether the extracellular binding and intracellular-domain-mediated cytoskeletal

reorganization functions can be uncoupled and recombined to generate new cell–cell connectivities, although previous studies indicate the potential for modularity[15,16,17,18,19].

Here we systematically explore the modularity of CAMs by fusing orthogonal extracellular binding domains (ECDs) to endogenous CAM intracellular domains (ICDs), thereby generating synthetic CAMs (synCAMs). We characterize the resulting cell–cell interfaces, and test whether synCAMs can program new multicellular organization.

Synthetic CAMs show native-like adhesion

We generated heterophilic synCAMs in which a well-characterized orthogonal binding interaction—the GFP–anti-GFP (nanobody) interaction—is fused to the ICDs of E-cadherin (ECAD), integrin β1 (ITGB1), integrin β2 (ITGB2), intercellular adhesion molecule 1 (ICAM-1), delta-like protein 1 (DLL1), junctional adhesion molecule B (JAM-B), neural cell adhesion molecule 1 (NCAM-1) and mucin 4 (MUC-4)[20] (Fig. 1b). The transmembrane region (TM) and ICD from the donor CAM was fused to the GFP or anti-GFP ECD.

We then tested whether cognate synCAM paired with symmetrically matched ICDs can drive junction formation between L929 mouse fibroblasts (a cell line with low endogenous adhesion that is used to assess cadherin differential adhesion sorting)[21,22]. Cells expressing cognate synCAMs were mixed in a flat-bottom, ultra-low-attachment (ULA) plate, and imaged using confocal microscopy (Fig. 1c). We compared synCAM-driven interfaces with those formed by native adhesion molecules (for example, wild-type (WT) ECAD) or by a simple tether (GFP or anti-GFP fused to a transmembrane domain lacking any ICD). synCAMs/tethers were expression matched (Extended Data Fig. 1).

Several synCAMs (ICDs: ECAD, ITGB1, ITGB2, ICAM-1 and MUC-4) formed extensive interfaces comparable to those observed with native cadherin. Native-like interfaces form despite these molecules completely lacking their large native extracellular domains. By contrast, the tether (no

ICD) did not form an extensive interface, showing only a small point of contact.

Several other synCAMs (ICDs: NCAM-1, JAM-B and DLL1) exhibited a distinct phenotype. The resulting interfaces were small, but with substantial interface enrichment of the GFP-labelled synCAMs (Fig. 1c). By contrast, the GFP signal in tethered cell pairs remained distributed throughout the entire membrane. Thus, these synCAMs drive a distinct phenotype of enriched spatial clustering at the interface when engaged.

To quantitatively analyse interface geometry for synCAM interactions (15–20 cell pairs), we measured contact angle—a standard metric of apparent cell–cell surface tension that is correlated with interface size[23,24,25] (Fig. 1d). We also measured the enrichment fraction (the fraction of GFP-tagged synCAM localized to the interface versus total membrane; Fig. 1e). These results show two main phenotypic classes of synCAMs: one class induces the formation of large, extensive cell–cell interfaces (ICDs: ECAD, ITGB1, ITGB2, and ICAM-1, MUC-4), and another class induces the formation of small but highly enriched interfaces (ICDs: NCAM-1, JAM-B and DLL1) (MUC-4 and ICAM-1 show hybrid behaviours). Each of these synCAM interface classes is distinct from the simple tether interaction.

The ICD determines interface strength

Both a strong ECD binding interaction and strong ICD coupling with the cytoskeleton could contribute to tight cell–cell interface formation. The synCAM modularity enables investigation of the relative ECD and ICD contributions to interface strength. Using the ICAM-1 synCAM as a testbed system, we characterized cell–cell interfaces with varied ECD affinity (using an affinity series of GFP nanobodies) or a deleted ICD[20] (Fig. 1f and Extended Data Fig. 1). Reducing the ECD affinity from a dissociation constant (K_d) of 0.7 nM to 3 μM (>10^3 fold) gradually decreases the resulting cell–cell contact angle, but even the weakest ECD exhibits a significantly expanded interface. By contrast, deletion of the ICAM-1 ICD, even in the presence of a high-affinity ECD, disrupts the interface completely. A similar modest decrease in cell–cell contact angle was observed for synCAMs with an ITGB1 ICD when the ECD K_d was varied

between 0.7 nM to 110 nM (Extended Data Fig. 2). These observations are consistent with a model in which cytomechanical changes mediated by the ICDs have a dominant role in determining the interface strength and morphology[23,24].

We also characterized how decreasing the ECD interaction affinity affects the interface enrichment phenotype of the NCAM-1 synCAM. The GFP receptor remains highly enriched at the interface even when the ECD affinity is varied over a range of K_d = 0.7 nM to 600 nM (Extended Data Fig. 2). Thus, the enriched interface phenotype also appears to be driven largely by the ICD identity.

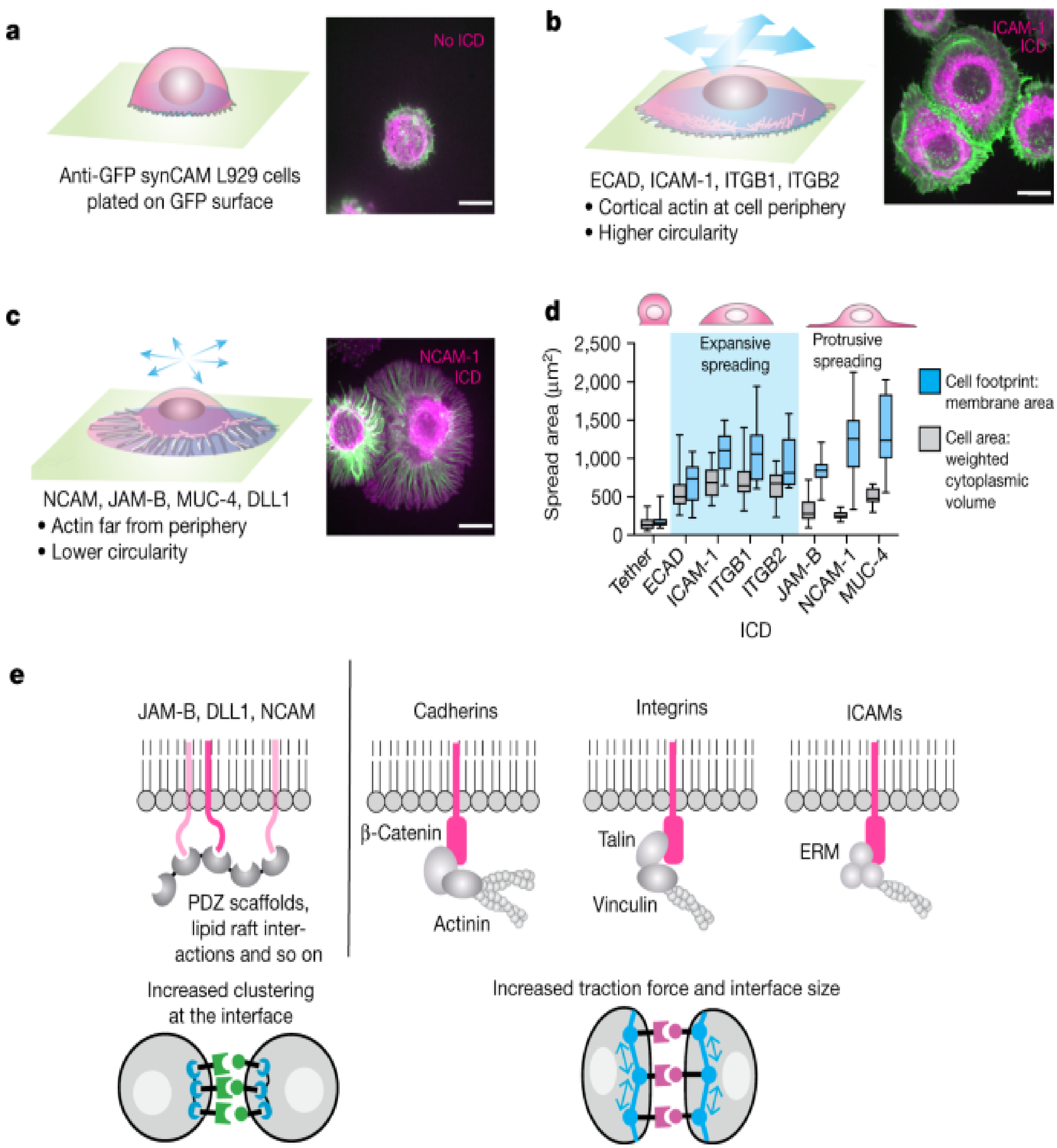
The dominance of the ICD over ECD affinity in determining adhesion properties was corroborated in competition sorting assays (Extended Data Fig. 3). Here, cells expressing two different anti-GFP synCAM (ICAM-1 ICD) variants compete to co-sort with GFP synCAM 'bait' cells. Higher-affinity cells preferentially sort with the bait cells to the core of the cell cluster. This complementary assay also indicates that the ICD primarily determines adhesion preferences. Expression of GFP–ICAM-1/anti-GFP–ICAM-1 at higher levels also increased the contact angle (Extended Data Fig. 4). By contrast, higher expression of the GFP/anti-GFP tethers did not change contact angle.

Two classes of interface morphologies

To examine synCAM interfaces in more detail, we used a more controlled assay in which an L929 cell expressing an anti-GFP synCAM interacts with a GFP-coated surface (Fig. 2 and Extended Data Fig. 5a). Here, because surface GFP is immobile and cannot rearrange, the interacting synCAM cells spread on the surface. After 75 min, cells were fixed and stained with Phalloidin to observe the actin cytoskeleton. A simple anti-GFP–tether interaction yielded minimal cell spreading on the GFP surface (Fig. 2a). However, synCAMs once again showed two distinct modes of spreading. Cells expressing synCAMs with ICDs from ICAM-1, ITGB1, ITGB2 and ECAD uniformly expanded on the GFP surface, developing a dense band of cortical actin along the cell periphery (Fig. 2b). Kinetic studies show that this larger spreading has a slow phase of tens of minutes to hours, consistent

with a requirement for cytoskeletal remodelling (Extended Data Fig. 5b–e). These synCAMs generate uniform 'expansive' spreading along the entire periphery of the cell. By contrast, synCAMs with the MUC-4, NCAM-1 and JAM-B ICDs yielded a 'fried egg' morphology—a smaller central cell mass was surrounded by thin membrane protrusions at the periphery (Fig. 2c). In these cases, lamellipodial and/or filopodial actin structures mediated radially 'protrusive' spreading. Overall, these surface-spreading studies are consistent with our previous cell–cell interface studies, as the expansive spreading synCAMs also lead to larger cell–cell interfaces and greater contact angles, whereas the protrusive spreading synCAMs formed small but highly enriched interfaces.

Fig. 2: SynCAM ICDs yield distinct mechanical and morphological properties.



a–**c**, Representative Phalloidin-stained images of L929 cells expressing the indicated synCAMs spreading on a GFP-coated surface. Scale bars, 10 μm. t = 2 h. Actin (Phalloidin stain) is shown in green; the full footprint of the cell (membrane label) is shown in purple. All of the images are shown at the same scale. **a**, The L929 cell expressing anti-GFP–tether (no ICD) shows minimal spreading. **b**, L929 cells expressing synCAMs with ICDs from ECAD, ICAM-1, ITGB1 and ITGB2 show an expansive spreading phenotype—the cell spreads in circular manner with cortical actin at the periphery of the cell footprint. See the spreading kinetic assays in Extended

Data Fig. 5. **c**, L929 cells expressing synCAMs with ICDs from NCAM-1, JAM-B and MUC-4 show a protrusive spreading phenotype (a 'fried egg' shape)—cortical actin does not spread very far, but the cell membrane footprint extends in a very thin layer beyond the cell, often with less circularity (that is, more filopodial or lamellopodial in nature). **d**, The full footprint of the cell (blue) and cell area (grey) for synCAM-mediated cell spreading. For the box plots, the centre line shows the median, the box limits show the 25th to 75th percentile and the whiskers show the minimum to maximum values. Cell area: $n = 23$ (tether), $n = 17$ (ECAD), $n = 23$ (JAM-B), $n = 16$ (ICAM-1), $n = 16$ (ITGB1), $n = 18$ (ITGB2), $n = 14$ (NCAM-1), $n = 12$ (MUC-4). Cell footprint: $n = 22$ (tether), $n = 21$ (ECAD), $n = 19$ (JAM-B), $n = 23$ (ICAM-1), $n = 16$ (ITGB1), $n = 12$ (ITGB2), $n = 14$ (NCAM-1), $n = 15$ (MUC-4). **e**, Known recruitment interactions of downstream intracellular proteins found in CAM ICDs. See the mutational analysis of ICD-binding motifs in Extended Data Fig.6.

Source data

We investigated how synCAM-driven cell spreading was altered by a series of small-molecule inhibitors of distinct actin regulators (Extended Data Fig. 6a). All synCAM-expressing cells displayed minimal spreading in the presence of latrunculin B, which disrupts actin filament formation, confirming the importance of cytoskeletal activity in all modes of cell spreading. By contrast, inhibiting contractility with blebbistatin (but still allowing actin polymerization) enabled synCAM cells to spread, but without controlled assembly of actin into distinct structures unique to the different synCAMs. This result emphasizes the competition between spreading and cortical contractility as a cell extends a new interface[24,26]. For protrusive spreading synCAMs (for example, JAM-B ICD), the lamellipodial sheets normally seen at the periphery of the cell are disrupted by CK666, indicating a role of its target, the ARP2/3 complex, in the formation of these thin protrusive structures.

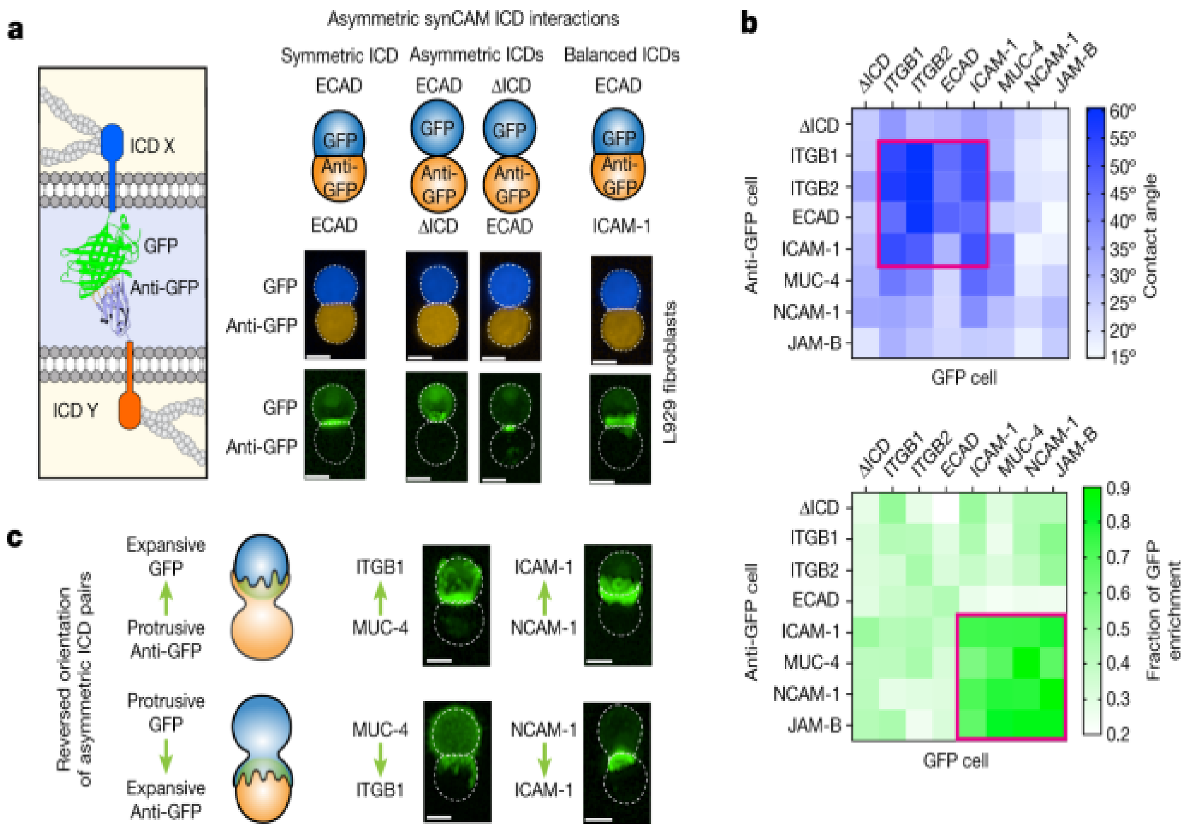
The distinct interface morphologies observed here can be explained by postulated mechanisms of the CAM ICDs (Fig. 2e). Although they individually differ in detail, the expansive ICDs (ECAD, ICAM-1, integrins) recruit adapter molecules such as β-catenin, talin, vinculin and ERM proteins, which are thought to engage the cortical actin cytoskeleton and

therefore drive expansion of the entire cell front[12,13,27]. By contrast, the protrusive ICDs (NCAM-1, JAM-B, DLL1) interact with PDZ scaffold proteins or lipid rafts, generally forming organized complexes that involve clustering or phase condensation[28,29,30,31]. The resulting spatially focused assemblies may then drive protrusive cytoskeletal responses such as the formation of filopodia and lamellipodia by recruiting and activating proteins such as N-WASP and ARP2/3. The importance of these ICD interaction domains in interface formation was confirmed by mutational analysis of key signalling motifs (Extended Data Fig. 6b–h).

Asymmetric interfaces

Many endogenous CAMs bind homophilically (for example, ECAD and JAM-B), yielding an interface with symmetric ICDs. However, many other endogenous CAMs participate in heterophilic interactions (for example, ITGB1, ITGB2 and ICAM-1), leading to cell–cell interfaces with different opposing ICDs. We therefore used the synCAM platform to investigate how symmetric versus asymmetric ICDs affect cell–cell interface morphology. We examined all of the possible pairs of different GFP–anti-GFP synCAMs in L929 fibroblasts (Fig. 3 and Extended Data Fig. 7).

Fig. 3: The balance of ICD properties determines asymmetric synCAM interface morphology.



a, Maximum projection of ×20 confocal microscopy images of pairwise synCAM interfaces ($t = 3$ h), showing symmetric ECAD ICDs (left), asymmetric ECAD and tether (ΔICD) interfaces (middle), and balanced asymmetric ECAD and ICAM-1 interfaces (right). Scale bars, 10 µm. The mCherry and BFP channels (top) and the GFP channels (bottom) of representative images from ten pairs over three independent replicates are shown. **b**, Quantification of the contact angle (top) and GFP enrichment (bottom) for pairwise asymmetric synCAM interfaces. $n = 10$. The combination of interfaces that exhibit the greatest contact angle or enrichment are outlined in red. **c**, Example ×20 confocal microscopy images of pairwise unbalanced asymmetric interfaces in which a protrusive synCAM binds to an expansive synCAM. $t = 3$ h. Scale bars, 10 µm. Representative images from ten pairs over three independent replicates are shown. Top, binding between a protrusive anti-GFP synCAM and an expansive GFP synCAM. Bottom, binding between a protrusive GFP synCAM and an expansive anti-GFP synCAM.

Source data

Asymmetric interfaces with a fully deleted ICD (tether) on one side of the interface exhibit significantly disrupted interfaces—they show minimal cell–cell interface expansion and contact angle increase (Fig. 3a,b). However, a large asymmetric interface can be formed if it pairs two expansive synCAMs (for example, ECAD–ICAM-1 or ECAD–ITGB2) (Fig. 3a,b). These findings suggest that large, expanded interfaces can form with asymmetric synCAMs if the opposing ICDs yield a balanced interaction. Analogously, asymmetric interfaces pairing two ICDs that both mediate GFP enrichment (for example, NCAM-1–MUC-4, NCAM-1–JAM-B) generate an interface enrichment phenotype that is similar to that of symmetric interfaces (Fig. 3a,b). Thus, to form a productive interface, the exact sequence of an opposing ICD is less critical than the presence of ICDs with matched strength and morphology.

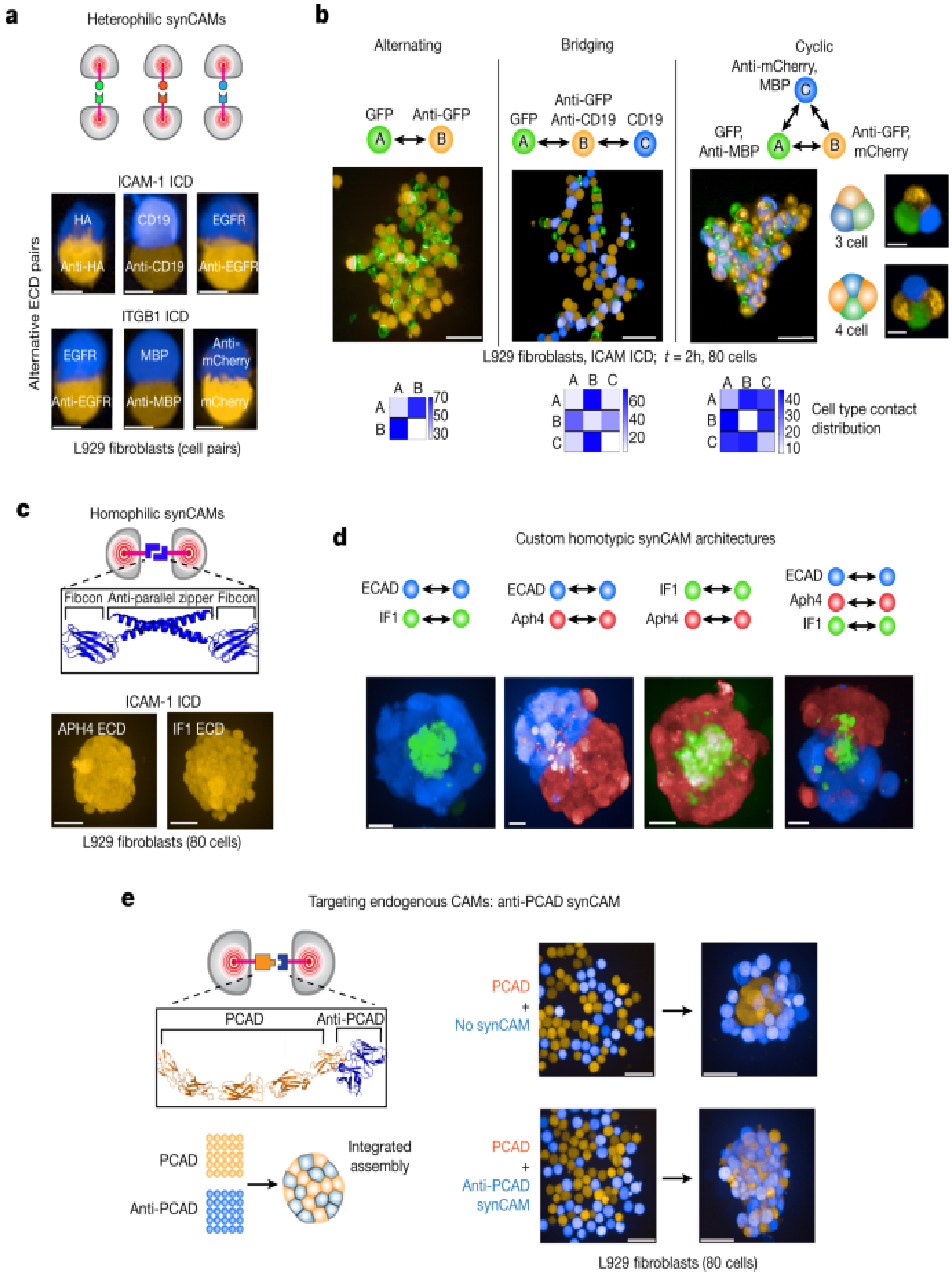
Notably, when we created heterotypic interfaces in which a cell with a protrusive synCAM binds to a cell with an expansive synCAM, the cells interacted with a consistent morphology—they formed an asymmetric interface in which the protrusive synCAM cell wraps around the expansive synCAM cell (Fig. 3c). These results show the diversity of interfaces that can be constructed with synCAM combinations.

Programming de novo cell assembly

Programming the formation of new multicellular tissues de novo requires dictating specific cellular connectivity within a multicellular system[5,32,33]. Previous efforts to orthogonally control multicellular assembly, both in bacteria and mammalian systems, have generally used surface-tethering approaches[5,32,33,34,35]. Notably, recent research has enabled custom patterning of engineered bacteria through the surface expression of orthogonal nanobody–antigen pairs[33]. Given the ability of synCAMs to direct cellular morphology and cytoskeletal structure, we tested whether synCAMs could be engineered with a wide range of orthogonal ECDs to also rationally program specific spatial connectivity. We found that functional synCAMs could be built with multiple distinct antibody–antigen binding pairs, including HA-tag–anti-HA single-chain variable fragment (scFv); maltose-binding protein (MBP)–anti-MBP nanobody; B cell surface antigen CD19–anti-CD19 scFv; tyrosine-protein kinase MET–anti-MET

nanobody; mCherry–anti-mCherry nanobody; and epidermal growth factor receptor (EGFR)–anti-EGFR nanobody (Fig. 4a and Supplementary Video 1). The orthogonality of distinct ECD synCAMs was confirmed using co-sorting assays, and we quantified their efficiency in excluding WT L929 cells from the multicellular assembly (Extended Data Fig. 8).

Fig. 4: Programming custom multicellular assemblies with homotypic and heterotypic synCAMs.



a, Heterophilic synCAMs with orthogonal extracellular recognition domains. Maximum projection of ×20 confocal microscopy cell–cell

interface images are shown of L929 cells expressing synCAMs with the indicated antibody–antigen pair ECDs and either ICAM-1 (top) or ITGB1 (bottom) TM/ICD. Scale bars, 10 µm. $t = 3$ h. Representative images are shown of four independent replicates. Experimental testing of orthogonal sorting is shown in Extended Data Fig. 8. See Supplementary Video 1 for a time-lapse analysis of the orthogonal assembly formation. **b**, Engineering of custom heterotypic assemblies. Maximum projection of ×20 confocal microscopy images of L929 cells expressing synCAMs with the indicated ECD recognition partners. $t = 2$ h. Assemblies form alternating (left), bridging (middle) and cyclic (right) patterning. Example images of isolated cyclic interactions are shown. $t = 2$ h. A time-lapse analysis is shown in Supplementary Video 2. Probability boxes of cell contact distribution are shown below. $n = 5$. Scale bars, 50 µm (left 3 images) and 10 µm (insets). **c**, synCAM design with a homophilic-binding leucine zipper ECD (top). Bottom, maximum projection of ×20 confocal microscopy images of L929 cells expressing homophilic-binding synCAMs with the Aph4 or IF1 leucine-zipper ECD and ICAM-1 TM/ICD (ULA round-bottom well, 80 cells total). Scale bars, 50 µm. $t = 24$ h. Representative images are shown of three independent replicates. **d**, ×20 confocal microscopy images of differential sorting between L929 cells expressing WT ECAD or the indicated homophilic-binding synCAMs. Scale bars, 20 µm. $t = 48$ h. Representative images are shown with additional independent replicates in Extended Data Fig. 9. $n = 15$ (ECAD–IF1), $n = 15$ (ECAD–Aph4), $n = 14$ (IF1–Aph4) and $n = 18$ (ECAD–IF1–Aph4). **e**, Schematic of the receptor design and differential sorting assay of L929 cells expressing WT PCAD (orange) and an anti-PCAD synCAM (anti-PCAD, blue) (left). The anti-PCAD synCAM contains an ICAM-1 TM/ICD. Right, maximum projection images of the sorting assay in which L929 cells expressing WT PCAD (orange) were mixed with parental (top) or synCAM (bottom) L929 cells (blue). Scale bars, 50 µm. $t = 0$ h and 24 h. Representative images are shown of four independent replicates with additional replicates shown in Extended Data Fig. 10.

Source data

We tested whether this set of orthogonal heterotypic synCAMs could program highly specific cell bonding patterns. (Fig. 4b and Supplementary Video 2). We constructed assemblies with the following patterns: (1) two

cell (A↔B) alternating heterophilic interactions (expression of a heterophilic GFP–anti-GFP synCAM pair in cells A and B); (2) three cell (A↔B↔C) bridging interactions (expression of orthogonal synCAMs in cells A and C, and both complementary synCAMs in the bridging cell B); (3) three cell (A↔B↔C↔A) cyclic interactions (expression of two orthogonal synCAMs in each of cells A, B and C). The resulting assemblies organize as dictated by the synCAM-defined cell–cell connectivities. Nearest neighbour distribution analysis (using Harmony image analysis software) showed that synCAM-specified interactions dominate assembly (Fig. 4b). In magnified images with low numbers of cells, the cyclic interaction set can lead to the predicted minimal 3 and 4 multicell assemblies (Fig. 4b). Thus, synCAM combinations can specify the precise bonding connectivities between cells.

We next engineered homotypic synCAMs from self-dimerizing coiled-coil ECD interactions. We used the Aph4 (a computationally designed leucine zipper[36]) and the IF1 (bovine ATPase inhibitor IF1) leucine zippers, as we anticipated that their antiparallel binding topologies might sterically favour intercellular *trans*-cell interactions over intracellular *cis* binding[36,37]. We also appended an intervening fibcon domain (a consensus FN3 domain from fibronectin) adjacent to the coiled-coil domains to provide additional separation from the juxtamembrane region, which could further favour *trans* cell interactions[38] (Fig. 4c).

We tested whether cells expressing orthogonal homophilic synCAMs could predictably generate structures with segregated compartments. Cells expressing three different orthogonal homotypic CAMs (WT ECAD, Aph4–ICAM-1 or IF1–ICAM-1) were mixed in combinations (Fig. 4d) and classified on the basis of the resulting assembly structures. The individual cell populations show clear sorting through their homophilic synCAMs, but most striking is the highly modular sorting behaviours that result. When cell types were mixed in a pairwise manner, we observed that the IF1 cells sort to the centre versus ECAD or Aph4. The ECAD and Aph4 cells sort into a two-lobed barbell structure. These relationships are maintained when all three cell types are mixed, yielding a structure with an ECAD–Aph4 barbell cell assembly with IF1 cells at the core (Extended Data Fig. 9 for assembly statistics). These results show how a toolkit of orthogonal synCAMs can

build multicompartment self-organizing structures with modularity and predictability.

Intercalation into native assemblies

We tested whether synCAMs could directly interface with a tissue held together by native adhesion molecules such as P-cadherin (PCAD). Thus, we engineered a synCAM with an anti-PCAD scFv fused to the ICAM-1 ICD (Fig. 4e, Extended Data Fig. 10 and Supplementary Video 3). These synthetic PCAD-targeting cells could effectively intercalate into a cell spheroid held together by PCAD. By contrast, cells lacking the synCAM were excluded and sorted to the exterior of the structure. Thus, synCAMs can integrate cells into assemblies formed by native adhesion molecules.

Use in primary and iPS cell-derived cells

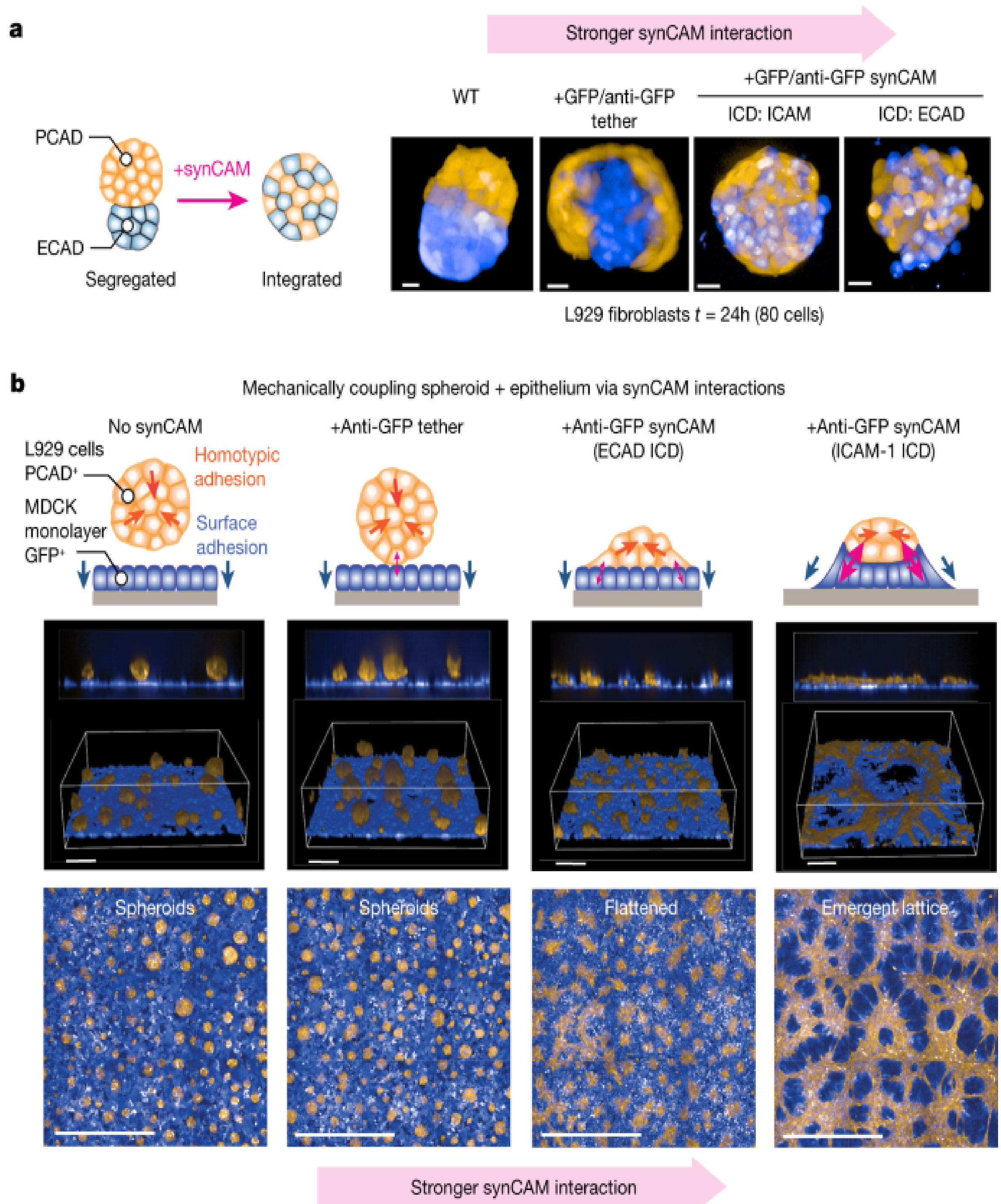
We tested whether synthetic adhesion molecules could function in primary cells and induced pluripotent stem (iPS) cell-derived cells. GFP/anti-GFP–ICAM-1 synCAMs and GFP/anti-GFP–tether molecules were expressed in several primary or iPS cell-derived cells (Extended Data Fig. 11). When ICAM-1-based synCAMs are expressed in primary human dermal fibroblasts, human mesenchymal stromal cells and iPS cell-derived smooth muscle cells, we observed strong localization of the GFP-tagged synCAMs to the interface formed with partner cells expressing a functional cognate anti-GFP synCAM. This synCAM relocalization to the heterotypic cell–cell interface was not observed either in unbound cells (GFP synCAM remains distributed throughout cell, not just at the interface) or when co-cultured with partner cells contained only an anti-GFP–tether (no ICD). These results demonstrate that synCAMs functionally engage each other in these different cell types in a manner that is dependent on cognate ECDs and the presence of functionally matched ICDs.

Remodelling tissue organization

We examined whether synthetic adhesion could remodel and reconfigure multicellular tissues organized by native CAMs. For example, L929 cells

expressing WT ECAD and WT PCAD differentially sort from each other into a bilobed assembly[6]. We examined whether introducing a GFP–anti-GFP synCAM interaction could force these two segregating populations to integrate (Fig. 5a). Expression of a heterotypic tether molecule converted the bilobed assembly into a two layered (core shell) structure, which maintains segregation, but slightly increases the number of heterophilic contacts relative to the bilobed assembly. By contrast, expression of the stronger synCAMs (ICAM-1 or ECAD ICDs) converted the bilobed structure into an integrated structure, with the two cell types in a single mixed compartment. These synCAMs could also force integration of differentially sorted L929 cell populations expressing WT PCAD or WT NCAD (Extended Data Fig. 12a,b). Thus, synCAMs can be used to systematically remodel multicell assemblies.

Fig. 5: Using synCAMs to reshape tissue organization.



a, Schematic of the experiment using synCAMs to force the integration of differentially sorting L929 populations. We start with L929 populations expressing WT ECAD (blue) or WT PCAD (orange), which leads to segregation into a binodal structure. The image shows how sorting is altered by the expression of integrating heterophilic synCAM with interactions of

different strengths (versus the tether receptor). Maximum projections of ×20 confocal microscopy images are shown. Scale bars, 20 μm. $t = 24$ h. See Supplementary Video 3 for a time-lapse analysis. A similar demonstration of synCAM integration of PCAD/NCAD-segregated cell populations is shown in Extended Data Fig. 7. **b**, L929 cells expressing WT PCAD (orange) mixed with an MDCK monolayer (blue) form spheroids that passively sit above the MCDK epithelial layer. Adding GFP–anti-GFP synCAMs with interactions of increasing strength (versus the tether receptor) leads to increasing mechanical coupling between the epithelial and spheroid tissues. When strong enough, the two cell types form a complex lattice-like network (ICAM synCAM). The images show assembly at $t = 24$ h. Both 3D magnified (top) and maximum projection zoomed-out (bottom) views are shown. Scale bars, 100 μm (top) and 1 mm (bottom). See Supplementary Video 4 for a time-lapse analysis of coupled tissue evolution.

To further examine tissue remodelling, we tested whether synCAMs could alter epithelial monolayers—a fundamental building block for diverse tissues and organs. For example, modulation of epithelial structure by interactions with mesenchymal cells is a common theme in development. We used Madin–Darby canine kidney (MDCK) cells as a starting epithelial cell layer. When a population of L929 cells expressing PCAD are added, they form segregated homotypic spheroid clusters that sit above the confluent MDCK epithelial layer. The starting epithelial (MDCK) and spheroid (PCAD-L929) tissues show minimal interactions, functioning as independent assemblies (Fig. 5b).

We examined whether introducing bridging synthetic adhesion interactions (using GFP–anti-GFP ECDs with symmetric ICDs) could force the distinct epithelial and spheroid tissues to interact. When a minimal tether interaction (no ICD) is added, the PCAD-L929 cells sit tightly on the MDCK epithelial layer, but still act independently, maintaining their segregated spheroid structure. However, introducing a stronger ECAD synCAM results in the PCAD-L929 spheroids spreading into flatter, aster-like bumps that more extensively contact the epithelial layer. Finally, adding the even stronger ICAM-1 synCAM bridging interaction causes substantial cooperative rearrangement of both tissues (Fig. 5b, Extended Data Fig. 12c and Supplementary Video 4). In this case, L929 cells organize into a continuous lattice network on top of the MDCK cells. Moreover, the MDCK epithelial

layer shows reduced confluence, perhaps because the strong bridging interaction between the L929 and MDCK cells appears to pull up MDCK cells from the surface in the intervening spaces of the lattice. We hypothesize that this cooperative tissue emerges from the opposing forces of the two tissues. The strong homotypic (PCAD) attraction among the L929 cells combined with the strong synthetic bridging interaction (synCAM) between the L929 cells and the MDCK cells results in these two populations adopting a mechanically balanced state. The resulting network is reminiscent of the self-organizing capillary tube network of activated endothelial cells. In brief, this lattice configuration appears to provide a solution that enables the L929 cells to simultaneously maintain a high degree of homotypic interaction, along with a high degree of heterotypic interaction with the MDCK epithelial layer. A similar emergent lattice network structure was observed in an analogous experiment in primary cells (primary mouse intestinal epithelial layer plus mouse embryonic fibroblasts; Extended Data Fig. 12d). In summary, synCAMs can systematically couple otherwise independent cell populations to yield multicell systems of which the cooperative mechanics yield complex tissue structures.

Discussion

Here we reveal the potential for engineering diverse synthetic adhesion molecules that share the design principles of native adhesion molecules, but that specify new and orthogonal connectivities between cells. Although metazoans deploy a plethora of CAMs to mediate diverse cellular interactions and tissue assembly, many more interfaces probably remain untapped by evolution. The synCAM design strategy used here integrates two mechanisms for controlling synthetic adhesion. First, the extracellular interaction domain specifies cell–cell connectivity (bonding), which can be either homophilic or heterophilic with precisely controlled affinity. Second, the intracellular domain dictates cytoskeletal reorganization and largely determines the interface mechanics and morphology. The orthogonality and tunability of extracellular domain recognition coupled to the modularity of intracellular domain output expands the possible set of interfaces that can be generated. This toolkit can therefore alter both cell–cell connectivity and the resulting interface type. Furthermore, mixing multiple synCAMs and native

CAMs to create a system of mechanically coupled cells can generate tissues with complex emergent structures.

The broad spectrum of adhesion ICDs amenable to chimeric engineering demonstrates that intracellular domain function is, to some degree, independent of the endogenous extracellular recognition mechanism. Notably, the simple extracellular interactions used here do not match the higher regulatory sophistication of many natural ECDs, which can also show *cis*-oligomerization, catch bonding and allosteric changes[8,39,40,41,42]. Nonetheless, synCAMs are still sufficient to assemble similar cell–cell interfaces. The modularity of CAMs provides insights into how many natural CAMs may have evolved. For example, proteins with cadherin ECDs are found in choanoflagellates (the closest single-cell relatives to metazoans), but they lack the metazoan ICDs[43,44]. These proteins may have been used by choanoflagellates to bind to food or substrates rather than for cell–cell adhesion, and then later co-opted for cell–cell adhesion through recombination with intracellular signalling domains[43].

This study supports a dominant role of the intracellular domain in dictating the character of CAM-mediated cell–cell interfaces. Tethering interactions between cells that do not engage the cytoskeleton are unable to generate strong, extensive interfaces, no matter what the extracellular binding affinity is. By contrast, synCAMs consisting of ICDs that engage the cytoskeleton facilitate a more complex morphology that depends on the identity of the ICD on each side of the interface. These observations are consistent with previous studies suggesting that cadherin ICDs remodel cortex tension to drive cell interface expansion and resistance to cell separation[23,24,45,46,47,48].

Finally, we show that synCAMs provide a versatile toolkit for programming multicellular structures, either de novo or by intercalating or remodelling tissues formed by native CAMs. The toolkit of synCAMs also enables systematic perturbation of self-organizing systems that could be used to analyse the mechanism of diverse developmental processes. In the future, these types of engineered adhesion molecules could potentially be applied to address therapeutic problems such as to precisely direct tissue repair and regeneration or to control the interactions and trafficking of immune and neural cells[49].

Methods

Materials

Oligonucleotides were purchased from Integrated DNA Technologies. In-Fusion cloning reagent, CloneAmp HiFi PCR Premix, the Lenti-X concentrator kit and Stellar chemically competent cells were purchased Takara Bio. Miniprep kits and spin columns were purchased from Qiagen. FuGENE HD Transfection Reagent was purchased from Promega. DMEM, GlutaMAX, Alexa Fluor 647 Phalloidin (A22287) and Alexa Fluor 555 Phalloidin (A34055) were purchased from Thermo Fisher Scientific. Fetal bovine serum (FBS) was purchased from the University of California, San Francisco (UCSF) Cell Culture Facility. L929 mouse fibroblast cells (ATCC, CCL-1) were purchased from the American Type Culture Collection. MDCK cells were a gift from the Mostov laboratory at UCSF. Primary dermal fibroblast cells (CC-2511), mouse embryonic fibroblasts (M-FB-481) and human bone-marrow-derived mesenchymal stem cells (PT-2501) were purchased from Lonza Bioscience. Nexcelom 3D 384-well ultra-low attachment treated round-bottom multi-well plates were purchased from Nexcelom Bioscience. Cellstar Cell-Repellent Surface 384-Well flat-bottom plates were purchased from Greiner Bio-One. The 384-well optical imaging flat clear-bottom TC-treated plates were purchased from Corning. H9 human pluripotent stem cells (hPSCs) (WA09) were purchased from WiCell. EDTA (46-034-CI) and growth-factor-reduced Matrigel (356231) were purchased from Corning (Corning). Geltrex, hESC-qualified (A1413302), Essential 8 Flex Medium Kit (A2858501), Essential 6 Flex Medium Kit (A1516401) and Advanced DMEM/F12 (12634028) were purchased from Thermo Fisher Scientific. Recombinant human/mouse/rat activin A protein (338-AC-050) was purchased from R&D Systems. FBS for iPS cells (1701) was purchased from ScienCell. CellMask deep red plasma membrane dye was purchased from Invitrogen. Phalloidin-iFluor 405 Reagent (ab176752) was purchased from Abcam.

The following antibodies were purchased and diluted in PBS before use according to the manufacturer's protocol: DYKDDDDK epitope tag Alexa Fluor 647-conjugated antibodies (1042E, rabbit, R&D Systems, IC8529R, AEOB0118081, 1:100); DYKDDDDK epitope tag Alexa Fluor 488-

conjugated antibodies (1042E, rabbit, R&D Systems, IC8529G, AEOA0521031, 1:100); anti-MYC-tag (9B11) mouse monoclonal antibodies (AlexaFluor 647 conjugate) (Cell Signaling Technology, 2233, 25, 1:100); anti-HA-tag (6E2) mouse monoclonal antibodies (AlexaFluor 647 conjugate) (Cell Signaling Technology, 3444, 15, 1:100); anti-human HGFR/c-MET (95106) AlexaFluor 488-conjugated antibodies (R&D Systems, FAB3582G, 1:50); anti-EGFR antibodies (DH8.3) (AlexaFluor 647) (Novusbio, 50599AF647, 1:50); and anti-6×His tag (HIS.H8) antibodies (Abcam, ab18184, 1:100).

Equipment

Cell sorting and flow cytometry was performed using the FACSAria II Cell Sorter or LSR II Flow Cytometer (Beckton-Dickinson). Confocal microscopy was performed using the Opera Phenix automated spinning-disk confocal microscope with a ×20 water-immersion objective in 384-well plates; the Nikon TiE with CSU-X1 spinning-disk confocal unit with ×60 and ×100 oil-immersion objectives; or the Zeiss LSM 980 with Airyscan 2 with a ×40 water-immersion objective.

Synthetic adhesion receptor construct design and cloning

All of the constructs were cloned into a pHR vector containing the SFFV promoter, Kozak consensus sequence and a cleavable signal sequence of influenza haemagglutinin (MKTIIALSYIFCLVFA)[50].

To design the synCAM constructs, transmembrane and intracellular regions from cellular adhesion molecules were identified from topology annotations in UniProt[51]. Codon-optimized genes encoding each CAM ICD and TM region were purchased from Integrated DNA Technologies and inserted into the vector using In-Fusion cloning. Each CAM TM and ICD region was fused to an extracellular binding domain (for example, GFP, anti-GFP) using In-Fusion cloning (Supplementary Table 2). Sequences for all nanobody or scFv ECDs were obtained from previously reported work or from publicly available patents[20,52,53,54,55,56]. For the experiments involving intestinal epithelial cells, an internal ribosome entry site (IRES) and a puromycin-*N*-acetyltransferase gene (Puro) were cloned downstream of the GFP–ICAM-1

and GFP–tether constructs within the pHR vector. Plasmids were sequence-verified by RF Biotech.

Lentivirus

Lentivirus was generated by cotransfecting vectors encoding packaging proteins (pMD2.G and p8.91) with the pHR plasmid of interest using the Fugene 6 HD transfection reagent (according to the manufacturer's protocol) in HEK293T cells plated in 6-well plates at approximately 70% confluence. Two days after transfection, viral supernatants were collected, passed through a 0.45 mm filter and used immediately for transduction.

For transduction of primary cells, lentivirus was concentrated 20-fold using the Lenti-X Concentrator kit (Takara) according to the manufacturer's protocol.

Cell lines

L929 and MDCK cells were cultured in DMEM containing 10% FBS. To generate stable cell lines, viral supernatant (50–400 μl) was diluted with 1.5 ml of medium and plated directly with cells (1×10^5 L929 or MDCK) in 12-well dishes. Then, 24 h after infection, the viral medium was replaced with normal growth medium and the cells were expanded into a T25 flask. The cells were stained for the appropriate epitope tag using a fluorescently tagged antibody and sorted for expression by fluorescence-activated cell sorting (FACS). Unless otherwise noted, a bulk-sorted population was used for each experiment. To generate the GFP–ICAM-1 and GFP–tether L929 cell lines with tuned expression levels, total virus added to the cells was titrated between 50 and 400 μl, and the cells were sorted for different synCAM expression levels using FACS. For the Aph4 and IF1 synCAMs, single-cell populations were established by sorting individual cells into a 96-well plate.

Antibody staining and flow cytometry analysis

To confirm the expression level of synCAMs in each cell line, the cells were analysed using FACS. The cells were detached with TrypLE and transferred

to a round-bottom 96-well plate. The cells were pelleted by centrifugation (for 4 min at 400*g*), the supernatant was removed and the cells were resuspended in 40 µl PBS containing a fluorescent-dye-conjugated antibody. Cells were stained for 50 min at 4 °C. The cells were then washed twice with PBS and resuspended in PBS with 5% FBS. The cells were then analysed using flow cytometry (BD LSR II, BD FACSDiva). The flow cytometry data were then analysed using FlowJo (TreeStar).

Contact angle and receptor enrichment measurements for cell–cell pairs

Before carrying out the experiment, all of the cell lines were detached using TrypLE, resuspended in 1 ml DMEM, counted and then diluted to 4×10^5 cells per ml. L929 cells stably expressing cytosolic BFP and a GFP synCAM were mixed 1:1 with L929 cells expressing cytosolic mCherry and an anti-GFP synCAM in a 384-well flat-bottom plate with a cell-repellent surface (3.2×10^4 cells, 80 µl total volume, 37 °C). At $t = 3$ h, the plates were imaged at ×20 magnification using fluorescence confocal microscopy (Phenix). Maximum-projection images were exported from the manufacturer's software (Harmony). Distinct cell pairs of similar size were identified, and contact angles were measured in ImageJ. The GFP enrichment percentage was determined in ImageJ by measuring the GFP signal localized at the cell–cell interface as a fraction of that present in the entire cell. Data analysis for the measured contact angle and enrichment values was performed in Prism 9 (GraphPad).

Cell spreading experiments

We characterized the rate, interface size and morphology of spreading synCAM cells on a GFP-coated surface. Purified GFP protein was diluted to a final concentration of 0.5 µM in PBS and enough volume was applied (~100 µl) to coat the bottom surface of an 8-well glass-bottom imaging chamber. This solution was incubated for 10 min on ice. Excess solution was removed and the chamber was rinsed with PBS. Next, the chamber was blocked with a solution of 10 mg ml^{-1} bovine serum albumin (BSA) and 1 mg ml^{-1} beta casein (Sigma-Aldrich) for a minimum of 1 h on ice. The blocking solution was removed and the chamber was washed three times

with PBS. When using CellVis (C8-1.5H-N) chambers, an anti-6×-His antibody (ab18184) and 6×-His-tagged GFP (ab134853) were used to obtain full coverage of the surface with GFP. A 100× dilution of antibody in PBS was incubated on the surface of the chamber for 1 h at 4 °C. After being washed three times with PBS, a 10 µg ml^{-1} solution containing His-tagged GFP was incubated on the surface for 1 h at 4 °C. Next, the chamber was blocked with a solution of 10 mg ml^{-1} BSA and 1 mg ml^{-1} beta casein (Sigma-Aldrich) for a minimum of 1 h on ice. The blocking solution was removed and the chamber washed three times with PBS.

To prepare the cells for the spreading assay, L929 cells were detached using Trypsin EDTA and resuspended in cell culture medium. Around 50 µl of resuspended cell solution from a confluent T25 flask was added to 200 µl of cell culture medium and placed into the imaging chamber. The chamber was then transferred to a spinning-disk confocal microscope equipped with an Oko Labs environmental control stage. Cells were imaged with a ×60 oil-immersion objective every 3 min over a period of 2 h. During the first 60–90 min, spreading of distinct cells onto the surface was observed by monitoring cytoplasmic fluorescent proteins expressed in the cytoplasm of the synCAM cells.

Images were analysed by binarizing the intensity to obtain a mask of the cell, which could then be used to calculate the total spread area (A) and perimeter (p) of the footprint. To characterize the morphology of the interface, the circularity ($c = p^2/4\pi A$) was calculated and compared between different synCAMs. These measurements were also performed using an anti-flag tag fluorescent antibody (labels synCAM constructs) to measure the area and morphology directly at the interface with the coverslip. To compare different spreading kinetics, the change in area over time was fitted with the following form: $A = bt^{1/4}$ where b is the spreading rate coefficient. This model was previously used to compare the kinetics of spreading cells on an adhesive surface[26]. Analysis was implemented in MATLAB (2020a).

Immunostaining

To visualize the actin cytoskeleton, spreading cells were fixed and stained for immunohistochemistry according to standard procedures. Cells were

fixed in 4% PFA in cytoskeleton buffer (10 mM PIPES, 100 mM NaCl, 300 mM sucrose, 1 mM EGTA, 1 mM $MgCl_2$) for 20 min on ice. Cells were then washed three times and permeabilized with 0.1% Triton X-100 solution in PBS for 10 min on ice and again washed three times. Cells were then blocked with 10% BSA in PBS (PBS-BSA) for a minimum of 1 h at 4 °C. To visualize the actin cytoskeleton, cells were stained with fluorescently labelled Phalloidin (conjugated with either Alexa 647, 555 or 405 fluorescent dyes). Cells were then imaged using a spinning-disk confocal microscope using a ×100 magnification objective. Cell peripheries were determined by staining with CellMask deep red plasma membrane stain (Invitrogen). For measurements investigating the effects of cytoskeletal inhibitors on cell spreading, cells were introduced into medium containing the inhibitor and allowed to spread on the GFP-coated surface (CK666 (100 µM), latrunculin B (5 µM), SMIFH2 (100 µM), blebbistatin (50 µM) inhibitors were purchased from Abcam). Cells were then fixed and stained according to the above procedure before being imaged using the Zeiss 980 Airyscan microscope and a ×40 water-immersion objective (Zen Blue).

Differential sorting assay

Before carrying out the experiment, all of the cell lines were detached using TrypLE, resuspended in 1 ml DMEM, counted and then diluted to 1×10^3 cells per ml. L929 cells stably expressing cytosolic BFP and an anti-GFP synCAM of varying affinity were mixed 1:1:1 with L929 cells expressing cytosolic mCherry and an anti-GFP synCAM of varying affinity, and L929 cells expressing GFP–ICAM-1 in distinct wells of a 384-well ULA round-bottom well (80 µl total volume). At $t = 24$ h, the wells were imaged at ×20 magnification using fluorescence confocal microscopy (Phenix).

Quantification of the sorting assay

To quantify the organization of different synCAM-expressing cells in the multicellular differential sorting assay, we calculated the radial distribution function $g(r)$ from multichannel 3D confocal stacks. Cells expressing mCherry and BFP were imaged at ×20 magnification with a z-step size of 10 µm. Each slice in the image stack was thresholded and binarized for each colour channel, and the centre of mass (COM) of the cluster was found. $g(r)$

was found by calculating the distance of each pixel from the COM and normalizing to the density of pixels within the cluster. To create a single value capturing the distribution of cells in the cluster, we calculated the COM of the $g(r)$ distribution and subtracted this value for the mCherry cells from the value for the BFP cells. Large values therefore indicate that mCherry cells are closer to the centre of the cluster and small values indicate that BFP cells are closer to the centre of the cluster. Image analysis was implemented in MATLAB (2020a).

Characterization of cell lines expressing orthogonal synCAMs

L929 cells stably expressing synCAMs with orthogonal heterophilic pairs and a cytosolic mCherry or BFP were generated. Before carrying out the experiment, cell lines were detached using TrypLE, resuspended in 1 ml DMEM, counted and then diluted to 4×10^5 cells per ml. Each pair was mixed 1:1 in a 384-well flat-bottom plate with a cell-repellent surface (3.2×10^4 cells, 80 µl total volume, 37 °C). At $t = 3$ h, the plates were imaged at ×20 magnification using fluorescence confocal microscopy (Phenix). Maximum-projection images were generated using the manufacturer's software.

To validate the orthogonality of the heterophilic synCAM pairs, a subset was characterized for the ability to differentially sort from parental L929 cells. The synCAM cell lines were detached using TrypLE, resuspended in 1 ml DMEM, counted and then diluted to 1×10^3 cells per ml. Parental L929 cells were detached using TrypLE, stained with CellTrace Far Red according to the manufacturer's instructions, and diluted to 1×10^3 cells per ml. Two synCAMs and the WT L929 cells were mixed 1:1:1 (80 µl total) in a ULA round-bottom well and imaged after 24 h at ×20 magnification using fluorescence confocal microscopy (Phenix). Maximum-projection images were then generated using the manufacturer's software (Harmony). Within the software, individual cells were segmented, and the centre of the assembly was calculated on the basis of the average position of all cells. The distance of the WT (Far Red) L929 cells and synCAM (BFP) cells from previously calculated centre of the assembly was then determined. The difference between the average distance of WT and synCAM cells was then

calculated and represented as a heat map, with greater distances corresponding to increased exclusion of WT cells from the assembly.

Design and characterization of cell lines expressing homotypic synCAMs

Homotypic synCAMs were designed to sterically impair ECD *cis*-interactions of the binding region. Antiparallel leucine zippers, which should favour *trans* over *cis* binding, were fused to a fibcon linker domain, which extends the receptor from the juxtamembrane region[37,38,36]. Efforts to design homotypic synCAMs without the fibcon linker were unsuccessful. These engineered ECDs were fused to an ICAM-1 TM/ICD.

L929 cells stably expressing the homophilic synCAM receptors and cytosolic mCherry were generated. Clonal cell lines were obtained through single-cell sorting. The cell lines were detached using TrypLE, resuspended in 1 ml DMEM, counted and then diluted to 1×10^3 cells per ml. The cells were incubated in a 384-well ULA round-bottom plate (80 cells, 80 µl total volume, 37 °C) for 24 h and then imaged by fluorescence confocal microscopy (Phenix). Maximum-projection images were generated using the manufacturer’s software (Harmony).

Targeting endogenous PCAD

L929 cells expressing WT PCAD and cytosolic mCherry were previously generated[6]. L929 cells expressing cytosolic BFP with or without stable expression of an anti-PCAD synCAM (ICAM-1 TM/ICD) were mixed 1:1 with L929 cells stably expressing WT PCAD and cytosolic mCherry in a 384-well ULA round-bottom plate (80 cells, 80 µl total volume, 37 °C) for 24 h and imaged using fluorescence confocal microscopy (Phenix). Maximum-projection images were generated using the manufacturer’s software (Harmony). Within the Harmony software, the total area encompassed by both the L929 cells expressing WT PCAD (mCherry) and the WT or anti-PCAD cells (BFP) was calculated for each maximum-projection image at each timepoint from distinct wells. The ratio of the area for BFP to mCherry cells was then calculated and plotted over 24 h, with an increased ratio corresponding to exclusion of BFP cells from the

multicellular assembly (Extended Data Fig. 10b). Moreover, for $t = 24$ h, the cells were segmented and the position of the centre of the assembly was calculated as the average position of the $mCherry^+$ and BFP^+ cells. The relative distance of the BFP^+ and $mCherry^+$ cells to the centre of the assembly was then calculated (Extended Data Fig. 10c) with a greater distance corresponding to increased exclusion of BFP^+ L929 cells.

Custom multicellular architecture

For the multicellular patterning experiments, L929 cell lines were detached using TrypLE, resuspended in 1 ml DMEM, counted and then diluted to 1×10^3 cells per ml. Before dilution, the Aph4 and IF1 synCAMs were stained with CellTrace Far Red and CFSE, respectively, according to the manufacturer's protocol.

Heterotypic assemblies

To generate the two-cell alternating pattern, L929 cells expressing GFP–ICAM-1 (cell 1) were mixed with L929 cells expressing cytosolic mCherry and LaG16–ICAM-1 (anti-GFP) (cell 2) (1:1 80 µl total). To generate the three-cell bridging pattern, L929 cells expressing GFP–ECAD (cell 1) were mixed with cells expressing cytosolic mCherry, LaG16–ECAD and anti-CD19–ICAM-1 (cell 2), and cells expressing cytosolic BFP and CD19–ICAM-1 (cell 3) (1:2:1 80 µl total). To generate the three-cell cyclic pattern, L929 cells expressing GFP–ECAD and anti-MBP–ICAM-1 (cell 1) were mixed with cells expressing LaG16–ECAD and mCherry–ICAM-1 (cell 2), and cells expressing MBP–ICAM-1, LaM4–ICAM-1 and cytosolic BFP (cell 3) (1:1:1 80 µl total). In all cases, the cells were plated in ULA round-bottom wells and imaged after 2 h using confocal microscopy (Phenix). Maximum-projection images from distinct wells were generated using the manufacturer's software (Harmony). To calculate the interaction probability tables, the cells were segmented in Harmony for each maximum-projection image. Cell–cell contacts were identified from the positions of the segmented cells, and the probability for each interaction was calculated and represented as a heat map.

To form the isolated three-cell and four-cell cyclic assemblies, L929 cells expressing GFP–ECAD and anti-MBP–ICAM-1 (cell 1); LaG16–ECAD and mCherry–ICAM-1 (cell 2); and MBP–ICAM-1, LaM4–ICAM-1 and cytosolic BFP (cell 3) were diluted to 4×10^3 cells per ml and plated in a flat-bottom well with a cell-repellent surface. Individual pairs were identified, and maximum-projection images were generated and exported.

Homotypic assemblies

L929 cells expressing Wt ECAD and cytosolic BFP, Aph4–ICAM-1 or IF1–ICAM-1 were mixed with each other (either individually or all three together) in ULA round-bottom wells (1:1 or 1:1:1, 80 µl total). The cells were imaged after 48 h using confocal microscopy. The maximum-projection images were generated from distinct wells using the manufacturer's software (Harmony) and were classified on the basis of assembly phenotype.

Primary cell culture

Adult human dermal fibroblasts (NHDF-Ad) and mouse embryonic fibroblasts (MEFs) were cultured in DMEM containing 10% FBS. Mesenchymal stem cells (MSCs) were cultured in mesenchymal stem cell growth medium (Lonza).

To generate stable cells expressing the synCAM constructs, viral supernatant (15 µl of ×20 concentrated virus) was diluted with 1.5 ml of medium and plated directly with cells grown to 80% confluency (5×10^4 MSCs, MEFs or NHDFs plated in a 12-well dish). Then, 24 h after transduction, the viral medium was replaced with normal growth medium and the cells were expanded into a 6-well dish. MEFs were further sorted for expression of synCAM constructs by FACS.

iPS cell-derived smooth muscle cells

Under the official approval from the UCSF Human Gamete, Embryo and Stem Cell Research Committee (GESCR) to F.F., we used the WA09 human embryonic stem cell lines purchased from WiCell in this study. These cell

lines and their original sample are completely de-identified and no authors had access to the identifiers.

hPSCs (WA09, WiCell) were maintained in E8 medium on Geltrex-coated six-well plates. Two days before initializing smooth muscle differentiation, hPSCs were dissociated with EDTA and replated into a Geltrex-coated six-well plate. Once hPSCs reached confluency, E8 medium was aspirated and replaced with 1 ml per well of Essential 6 with 100 ng ml^{-1} activin A. The next day, the medium was aspirated and replaced with 2 ml per well of E6 medium with 10 ng ml^{-1} BMP4. Two days later, the medium was aspirated and replaced with 2 ml per well of E6 medium with 10 ng ml^{-1}BMP4. For days 5–9, cells were maintained with fresh E6 medium + 2% FBS every other day. From day 10 onwards, the medium was replaced three times per week with Advanced DMEM/F12 + 10% FBS.

To generate SMCs with stable expression of synCAMs, the SMCs were grown to 80% confluency in a 96-well plate and transduced with 1 µl of 20× concentrated virus. After 24 h, the medium was removed and replaced with fresh medium.

Mouse intestinal epithelial cells

Intestinal epithelium was isolated and cultured as previously described[57]. In brief, small intestinal crypts were dissociated from the duodenum of male C57BL/6 mice aged 6–12 weeks. The tissue was placed into ice-cold PBS with 15 mM EDTA for 30 min, then vortexed vigorously in multiple fractions to release crypts. The supernatant containing the crypts was filtered on a 70 µM mesh, and then the crypts were pelleted and resuspended in growth-factor-reduced Matrigel and cultured as 3D enteroids with ENR medium (Advanced DMEM/F12 (Thermo Fisher Scientific, 12634-028) with 1× N2 (Thermo Fisher Scientific, 17502-048), 1× B27 (Thermo Fisher Scientific, 17504-044), 10 mM HEPES (Thermo Fisher Scientific, 15630080), 1× GlutaMAX (Thermo Fisher Scientific, 35050-061), 1 mM *N*-acetylcysteine (Sigma-Aldrich A9165), 100 U ml^{-1} penicillin and 100 mg ml^{-1} streptomycin (Corning, 30-002), supplemented with 50 ng ml^{-1} EGF (Sigma Aldrich, E9644.2MG), 100 ng ml^{-1} Noggin (R&D 6057-NG/CF),

and 5% R-spondin-conditioned medium). The medium was changed every 3 days and organoids were mechanically dissociated and passaged weekly.

For these experiments, mice were maintained in the University of California San Francisco (UCSF) specific pathogen-free animal facility. All maintenance and experiments were performed in accordance with the guidelines established by the Institutional Animal Care and Use Committee and Laboratory Animal Resource Center. All experimental procedures were approved by the Laboratory Animal Resource Center at UCSF. Mice were housed in the UCSF LARC Animal Care Facilities at UCSF Parnassus. They were housed in an individual specific pathogen-free suite. They were housed with up to five mice per cage in ventilator cages, with ad libitum food and water under a 12 h–12 h light–dark cycle and controlled temperature and humidity conditions (20–26 °C and 30–70%).

For expression of synCAM constructs, organoids were transduced with lentivirus as previously described[58]. First, 3D enteroids were dissociated into single cells using TrypLe, which were then grown in growth-factor-reduced Matrigel and transduction medium (NR medium supplemented with 50% Wnt3a-conditioned medium, 10 μM nicotinamide (Sigma-Aldrich, N3376-100G), 5 μM CHIR (Sigma-Aldrich, SML1046-5MG) and 10 μM Y-27632 (Sigma-Aldrich, Y0503-1MG)) for 3–5 days to enrich for stem cells. Enteroids were then dissociated, pelleted, resuspended in transduction medium containing 8 μg ml^{-1} polybrene (Sigma-Aldrich, H9268-5G) and concentrated lentivirus, centrifuged at 600*g* for 1 h at 32 °C, then incubated at 37 °C for 6 h. Cells were then pelleted and resuspended in Matrigel and grown in transduction medium for 3 days, then switched to ENR medium. After amplification, antibiotic selection was performed by adding 1 μg ml^{-1} puromycin (Thermo Fisher Scientific, A1113803) to the medium.

Primary cell–cell adhesion assays

GFP–ICAM-1, GFP–tether, anti-GFP–fibcon–ICAM-1 or anti-GFP–fibcon–tether were transduced in MSCs, NHDFs or SMCs. For these experiments, a fibcon linker domain was included for both the anti-GFP–ICAM-1 and anti-GFP–tether constructs to improve expression in primary cells. All GFP-expressing cells were co-transduced with a plasmid for expression of

cytosolic BFP, and all anti-GFP expressing cells were co-transduced with a construct expressing cytosolic mCherry. Then, 24 h after transduction, the medium was removed and replaced with fresh medium. After 4 to 7 days, the MSCs, SMCs or NHDFs were detached with TryplE, resuspended in medium and plated in a 384-well plate. Then, 24 h after plating, the wells were imaged using fluorescence confocal microscopy (Phenix).

Modifying 3D architecture

L929 cells stably expressing WT PCAD, cytosolic mCherry and LaG16–synCAM (ICAM-1, ECAD or tether control) were mixed 1:1 with L929 cells stably expressing WT ECAD, cytosolic BFP and a GFP–synCAM (ICAM-1, ECAD or tether control) in a ULA round-bottom plate (80 total cells, 80 µl, 24 h, 37 °C). Before mixing, the L929 cell lines were detached using TrypLE, resuspended in 1 ml DMEM, counted and then diluted to 1×10^3 cells per ml. The assemblies were imaged using fluorescence confocal microscopy (Phenix, ×20 magnification), and maximum-projection images from distinct wells were generated using the manufacturer's software and are shown.

To modify the assembly between L929 cells expressing WT NCAD and L929 cells expressing WT PCAD, the experiment was performed exactly as described above with L929 cells expressing WT NCAD and cytosolic GFP in place of the WT ECAD cells.

Modifying 2D layering

An adherent layer of MDCK cells expressing cytosolic BFP and GFP–Tether, GFP–ICAM-1 or GFP–ECAD was formed within wells of a 384-well plate (16,000 cells plated per well). After 48 h, L929 cells expressing WT PCAD, cytosolic mCherry, and LaG16–ICAM-1, LaG16–tether, LaG16–ECAD or no additional receptor were added (24,000 cells per well). The interaction between the two layers was imaged using fluorescence confocal microscopy (Phenix) for 24 h. The zoomed-out images of the assemblies were formed by stitching together nine adjacent fields of view after exporting the images from the manufacturer's software. Both the

roundness and surface area of the mCherry$^+$ assembly was quantified for each field of the experiment within the manufacturer's software (Harmony).

Modifying 2D layering on intestinal epithelial organoids

Monolayer enteroid cultures were established as previously described[59]. 3D Enteroids were dissociated into single cells using TrypLE, washed in PBS and stained using CellTrace. A total of 150,000 cells expressing either GFP–ICAM-1 or GFP–Tether were plated onto a 384-well pate precoated with 5% growth-factor-reduced Matrigel in 40 µl ENR medium supplemented with 3 µM CHIR and 10 µM Y-27632. After 4 h, an additional 60 µl of ENR medium was added to each well. Then, 24 h after plating, the enteroid monolayers, mouse embryonic fibroblast cells (MEFs) expressing anti-GFP–fibcon–tether or anti-GFP–fibcon–ICAM-1 and cytosolic mCherry were added (16,000 cells). After 24 h, the wells were imaged using fluorescence confocal microscopy (Phenix). Maximum-projection images and 3D images were exported from the manufacturer's software (Harmony).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Experimental data supporting the conclusions of this study are available within the Article and the Supplementary Information. All databases used in this study are publicly available. For identifying protein sequences and domain architecture, the Universal Protein Resource (https://www.uniprot.org/) was used. For the identification of linear motifs within CAM ICDs, the Eukaryotic Linear Motif (ELM) resource (http://elm.eu.org/) was used. Additional microscopy replicates are available at Figshare (https://doi.org/10.6084/m9.figshare.21647546.v1). Source data are provided with this paper.

References

1. Honig, B. & Shapiro, L. Adhesion protein structure, molecular affinities, and principles of cell-cell recognition. *Cell* **181**, 520–535 (2020).

2. Ley, K., Laudanna, C., Cybulsky, M. I. & Nourshargh, S. Getting to the site of inflammation: the leukocyte adhesion cascade updated. *Nat. Rev. Immunol.* **7**, 678–689 (2007).

3. Sanes, J. R. & Zipursky, S. L. Synaptic specificity, recognition molecules, and assembly of neural circuits. *Cell* **181**, 536–556 (2020).

4. Collinet, C. & Lecuit, T. Programmed and self-organized flow of information during morphogenesis. *Nat. Rev. Mol. Cell Biol.* **22**, 245–265 (2021).

5. Todhunter, M. E. et al. Programmed synthesis of three-dimensional tissues. *Nat. Methods* **12**, 975–981 (2015).

6. Toda, S., Blauch, L. R., Tang, S. K. Y., Morsut, L. & Lim, W. A. Programming self-organizing multicellular structures with synthetic cell-cell signaling. *Science* **361**, 156–162 (2018).

7. Cavallaro, U. & Dejana, E. Adhesion molecule signalling: not always a sticky business. *Nat. Rev. Mol. Cell Biol.* **12**, 189–197 (2011).

8. Rubinstein, R. et al. Molecular logic of neuronal self-recognition through protocadherin domain interactions. *Cell* **163**, 629–642 (2015).

9. Tsai, T. Y.-C., Garner, R. M. & Megason, S. G. Adhesion-based self-organization in tissue patterning. *Annu. Rev. Cell Dev. Biol.* **38**, 349–374 (2022).

10. Lecuit, T. & Lenne, P.-F. Cell surface mechanics and the control of cell shape, tissue patterns and morphogenesis. *Nat. Rev. Mol. Cell Biol.* **8**, 633–644 (2007).

11. Arslan, F. N., Eckert, J., Schmidt, T. & Heisenberg, C.-P. Holding it together: when cadherin meets cadherin. *Biophys. J.* **120**, 4182–4192 (2021).

12. Kinashi, T. Intracellular signalling controlling integrin activation in lymphocytes. *Nat. Rev. Immunol.* **5**, 546–559 (2005).

13. Yap, A. S. & Kovacs, E. M. Direct cadherin-activated cell signaling: a view from the plasma membrane. *J. Cell Biol.* **160**, 11–16 (2003).

14. Parsons, J. T., Horwitz, A. R. & Schwartz, M. A. Cell adhesion: integrating cytoskeletal dynamics and cellular tension. *Nat. Rev. Mol. Cell Biol.* **11**, 633–643 (2010).

15. Geiger, B., Salomon, D., Takeichi, M. & Hynes, R. O. A chimeric N-cadherin/β_1-integrin receptor which localizes to both cell-cell and cell-matrix adhesions. *J. Cell Sci.* **103**, 943–951 (1992).

16. LaFlamme, S. E., Thomas, L. A., Yamada, S. S. & Yamada, K. M. Single subunit chimeric integrins as mimics and inhibitors of endogenous integrin functions in receptor localization, cell spreading and migration, and matrix assembly. *J. Cell Biol.* **126**, 1287–1298 (1994).

17. Lilienbaum, A., Reszka, A. A., Horwitz, A. F. & Holt, C. E. Chimeric integrins expressed in retinal ganglion cells impair process outgrowth in vivo. *Mol. Cell. Neurosci.* **6**, 139–152 (1995).

18. Young, B. A. et al. The cytoplasmic domain of the integrin alpha9 subunit requires the adaptor protein paxillin to inhibit cell spreading but promotes cell migration in a paxillin-independent manner. *Mol. Biol. Cell* **12**, 3214–3225 (2001).

19. Schoenit, A. et al. Tuning epithelial cell-cell adhesion and collective dynamics with functional DNA–E-cadherin hybrid linkers. *Nano Lett.* **22**, 302–310 (2022).

20. Fridy, P. C. et al. A robust pipeline for rapid production of versatile nanobody repertoires. *Nat. Methods* **11**, 1253–1260 (2014).

21. Nose, A., Nagafuchi, A. & Takeichi, M. Expressed recombinant cadherins mediate cell sorting in model systems. *Cell* **54**, 993–1001

(1988).

22. Foty, R. A. & Steinberg, M. S. The differential adhesion hypothesis: a direct evaluation. *Dev. Biol.* **278**, 255–263 (2005).

23. Maître, J.-L. et al. Adhesion functions in cell sorting by mechanically coupling the cortices of adhering cells. *Science* **338**, 253–256 (2012).

24. Winklbauer, R. Cell adhesion strength from cortical tension—an integration of concepts. *J. Cell Sci.* **128**, 3687–3693 (2015).

25. Cerchiari, A. E. et al. A strategy for tissue self-organization that is robust to cellular heterogeneity and plasticity. *Proc. Natl Acad. Sci. USA* **112**, 2287–2292 (2015).

26. Cuvelier, D. et al. The universal dynamics of cell spreading. *Curr. Biol.* **17**, 694–699 (2007).

27. Barreiro, O. et al. Dynamic interaction of VCAM-1 and ICAM-1 with moesin and ezrin in a novel endothelial docking structure for adherent leukocytes. *J. Cell Biol.* **157**, 1233–1245 (2002).

28. Beutel, O., Maraspini, R., Pombo-García, K., Martin-Lemaitre, C. & Honigmann, A. Phase separation of zonula occludens proteins drives formation of tight junctions. *Cell* **179**, 923–936 (2019).

29. Sytnyk, V., Leshchyns'ka, I., Nikonenko, A. G. & Schachner, M. NCAM promotes assembly and activity-dependent remodeling of the postsynaptic signaling complex. *J. Cell Biol.* **174**, 1071–1085 (2006).

30. Tetzlaff, F. et al. MPDZ promotes DLL4-induced Notch signaling during angiogenesis. *eLife* **7**, e32860 (2018).

31. Schwayer, C. et al. Mechanosensation of tight junctions depends on ZO-1 phase separation and flow. *Cell* **179**, 937–952.e18 (2019).

32. Gartner, Z. J. & Bertozzi, C. R. Programmed assembly of 3-dimensional microtissues with defined cellular connectivity. *Proc. Natl Acad. Sci. USA* **106**, 4606–4610 (2009).

33. Glass, D. S. & Riedel-Kruse, I. H. A synthetic bacterial cell-cell adhesion toolbox for programming multicellular morphologies and patterns. *Cell* **174**, 649–658 (2018).

34. Chao, G. et al. helixCAM: a platform for programmable cellular assembly in bacteria and human cells. *Cell* **185**, 3551–3567 (2022).

35. Kim, H. et al. 4-Bit adhesion logic enables universal multicellular interface patterning. *Nature* **608**, 324–329 (2022).

36. Negron, C. & Keating, A. E. A set of computationally designed orthogonal antiparallel homodimers that expands the synthetic coiled-coil toolkit. *J. Am. Chem. Soc.* **136**, 16544–16556 (2014).

37. Rhys, G. G. et al. Maintaining and breaking symmetry in homomeric coiled-coil assemblies. *Nat. Commun.* **9**, 4132 (2018).

38. Jacobs, S. A. et al. Design of novel FN3 domains with high stability by a consensus sequence approach. *Protein Eng. Des. Sel.* **25**, 107–117 (2012).

39. Hynes, R. O. Integrins: bidirectional, allosteric signaling machines. *Cell* **110**, 673–687 (2002).

40. Luo, B.-H. & Springer, T. A. Integrin structures and conformational signaling. *Curr. Opin. Cell Biol.* **18**, 579–586 (2006).

41. Wu, Y. et al. Cooperativity between *trans* and *cis* interactions in cadherin-mediated junction formation. *Proc. Natl Acad. Sci. USA* **107**, 17592–17597 (2010).

42. Brasch, J., Harrison, O. J., Honig, B. & Shapiro, L. Thinking outside the cell: how cadherins drive adhesion. *Trends Cell Biol.* **22**, 299–310 (2012).

43. Abedin, M. & King, N. The premetazoan ancestry of cadherins. *Science* **319**, 946–948 (2008).

44. King, N., Hittinger, C. T. & Carroll, S. B. Evolution of key cell signaling and adhesion protein families predates animal origins. *Science* **301**, 361–363 (2003).

45. Lecuit, T. & Yap, A. S. E-cadherin junctions as active mechanical integrators in tissue dynamics. *Nat. Cell Biol.* **17**, 533–539 (2015).

46. Manning, M. L., Foty, R. A., Steinberg, M. S. & Schoetz, E.-M. Coaction of intercellular adhesion and cortical tension specifies tissue surface tension. *Proc. Natl Acad. Sci. USA* **107**, 12517–12522 (2010).

47. Krieg, M. et al. Tensile forces govern germ-layer organization in zebrafish. *Nat. Cell Biol.* **10**, 429–436 (2008).

48. Amack, J. D. & Manning, M. L. Knowing the boundaries: extending the differential adhesion hypothesis in embryonic cell sorting. *Science* **338**, 212–215 (2012).

49. Lim, W. A. The emerging era of cell engineering: harnessing the modularity of cells to program complex biological function. *Science* **378**, 848–852 (2022).

50. Guan, X. M., Kobilka, T. S. & Kobilka, B. K. Enhancement of membrane insertion and function in a type IIIb membrane protein following introduction of a cleavable signal peptide. *J. Biol. Chem.* **267**, 21995–21998 (1992).

51. UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).

52. Zimmermann, I. et al. Synthetic single domain antibodies for the conformational trapping of membrane proteins. *eLife* **7**, e34317 (2018).

53. Zhao, N. et al. A genetically encoded probe for imaging nascent and mature HA-tagged proteins in vivo. *Nat. Commun.* **10**, 2947 (2019).

54. Grupp, S. A. et al. Chimeric antigen receptor-modified T cells for acute lymphoid leukemia. *N. Engl. J. Med.* **368**, 1509–1518 (2013).

55. Roovers, R. C. et al. A biparatopic anti-EGFR nanobody efficiently inhibits solid tumour growth. *Int. J. Cancer* **129**, 2013–2024 (2011).

56. Williams, J. Z. et al. Precise T cell recognition programs designed by transcriptionally linking multiple receptors. *Science* **370**, 1099–1104 (2020).

57. Mahe, M. M. et al. Establishment of gastrointestinal epithelial organoids. *Curr. Protoc. Mouse Biol.* **3**, 217–240 (2013).

58. Andersson-Rolf, A., Fink, J., Mustata, R. C. & Koo, B.-K. A video protocol of retroviral infection in primary intestinal organoid culture. *J. Vis. Exp.* **90**, e51765 (2014).

59. Thorne, C. A. et al. Enteroid monolayers reveal an autonomous WNT and BMP circuit controlling intestinal epithelial growth and organization. *Dev. Cell* **44**, 624–633 (2018).

60. Celli, L., Ryckewaert, J.-J., Delachanal, E. & Duperray, A. Evidence of a functional role for interaction between ICAM-1 and nonmuscle alpha-actinins in leukocyte diapedesis. *J. Immunol.* **177**, 4113–4121 (2006).

61. Anthis, N. J. et al. The structure of an integrin/talin complex reveals the basis of inside-out signal transduction. *EMBO J.* **28**, 3623–3632 (2009).

62. García-Alvarez, B. et al. Structural determinants of integrin recognition by talin. *Mol. Cell* **11**, 49–58 (2003).

63. Huber, A. H. & Weis, W. I. The structure of the β-catenin/E-cadherin complex and the molecular basis of diverse ligand recognition by β-catenin. *Cell* **105**, 391–402 (2001).

64. Little, E. B., Edelman, G. M. & Cunningham, B. A. Palmitoylation of the cytoplasmic domain of the neural cell adhesion molecule N-CAM serves as an anchor to cellular membranes. *Cell Adhes. Commun.* **6**, 415–430 (1998).

Acknowledgements

We thank D. Mullins, Z. Gartner, V. Weaver, M. Kutys and the members of the Lim laboratory and Cell Design Institute for discussions, assistance and advice. This work was supported by a NSF Center for cellular construction grant (DBI-1548297), the NIH NCI (U01CA265697) and the UCSF Cell Design Institute. A.J.S. is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation (DRG, 2355-19) and is a Hartz Cell Design Fellow. A.R.H was supported by an NSERC Discovery Grant (RGPIN-2022-04933). J.G. was supported by the National Institute of Neurological Disorders and Stroke, through UCSF grant number 2R25NS070680-11.

Author information

Author notes

1. Andrew R. Harris

   Present address: Department of Mechanical and Aerospace Engineering, Carleton University, Ottawa, Ontario, Canada

2. Wesley L. McKeithan

   Present address: Maze Therapeutics, San Francisco, CA, USA

Authors and Affiliations

1. UCSF Cell Design Institute, University of California, San Francisco, CA, USA

   Adam J. Stevens, Josiah Gerdts, Ki H. Kim, Wesley L. McKeithan & Wendell A. Lim

2. Department of Cellular and Molecular Pharmacology, University of California, San Francisco, CA, USA

   Adam J. Stevens, Josiah Gerdts, Ki H. Kim, Jonathan T. Ramirez, Wesley L. McKeithan, Faranak Fattahi & Wendell A. Lim

3. Center for Cellular Construction, University of California, San Francisco, CA, USA

   Adam J. Stevens, Andrew R. Harris, Josiah Gerdts, Ki H. Kim, Wesley L. McKeithan, Daniel A. Fletcher & Wendell A. Lim

4. Department of Bioengineering, University of California, Berkeley, CA, USA

   Andrew R. Harris & Daniel A. Fletcher

5. Department of Neurology, Weill Institute for Neuroscience, University of California, San Francisco, CA, USA

   Josiah Gerdts

6. Program in Craniofacial Biology, University of California, San Francisco, CA, USA

   Coralie Trentesaux & Ophir D. Klein

7. Department of Orofacial Sciences, University of California, San Francisco, CA, USA

   Coralie Trentesaux & Ophir D. Klein

8. Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California, San Francisco, CA, USA

   Jonathan T. Ramirez & Faranak Fattahi

9. Department of Pediatrics, Cedars-Sinai Medical Center, Los Angeles, CA, USA

   Ophir D. Klein

10. Chan Zuckerberg Biohub, San Francisco, CA, USA

   Daniel A. Fletcher

Contributions

A.J.S., A.R.H., C.T., F.F., O.D.K., D.A.F. and W.A.L. designed research. A.J.S., J.G., K.H.K. and W.L.M. cloned plasmids and generated cell lines. A.J.S. performed cell–cell adhesion experiments. A.R.H. performed cell-spreading adhesion experiments. A.J.S. C.T., J.T.R. and K.H.K. performed adhesion experiments in primary and iPS cell-derived cells. A.J.S. and A.R.H. analysed data. A.J.S., A.R.H., D.A.F. and W.A.L. wrote the paper.

Corresponding author

Correspondence to Wendell A. Lim.

Ethics declarations

Competing interests

W.A.L. is an advisor for Allogene and SciFi Foods, and owns equity in Gilead and Intellia. A patent application has been filed by the University of California San Francisco in relation to the engineered adhesion molecules reported in this work with W.A.L. and A.J.S. listed as inventors (PCT/US2021/057601). The other authors declare no competing interests.

Peer review

Peer review information

Nature thanks Matthias Lutolf and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Characterization of synCAM expression and function in L929 fibroblast cells (*linked to main* Fig. 1c–f).

(**a**) FACS analysis of GFP (left) and αGFP (right) synCAM expression in L929 fibroblast cells following cell sorting. Surface expression of each synCAM is measured using labelled anti-FLAG tag antibody. The CAM TM and ICD domain for each construct is indicated (tether = control lacking ICD, DLL1 = Delta-like Protein 1, JAM-B = Junction Adhesion Molecule B, NCAM-1 = Neural Cell Adhesion molecule 1, MUC-4 = Mucin 4, ICAM-1 = Intercellular Adhesion Molecule 1, Ecad = E-cadherin, Intβ1 = beta 1 integrin, Intβ2 = beta 2 integrin). Analysis shows that surface expression levels of the tether and alternative synCAM constructs are well matched. (**b**) Additional replicates of synCAM cell–cell adhesion interface analysis. Maximum projection of 20x confocal microscopy images of pairwise synCAM interfaces (t = 3 h): GFP-expressing cell (blue) is bound to an αGFP expressing cell (orange). The GFP channel of the interfaces is shown, highlighting differences of receptor enrichment. Four out of twenty additional examples are shown here. (**c**) FACS analysis of αGFP synCAM expression in L929 fibroblast cells expressing cytosolic mCherry (left) or BFP (right) following cell sorting. The CAM TM and ICD domain for each construct is ICAM-1, and the GFP-binding llama nanobody (LaG) ECD for each construct is indicated. This analysis shows that this series of alternative affinity synCAMs are expressed at comparable levels. (**d**) Maximum projection of 20X confocal microscopy images of pairwise synCAM interfaces (t = 3 h): GFP-expressing cell (blue) is bound to an αGFP expressing cell with the indicated binding K_d (orange).

Extended Data Fig. 2 Changing ECD affinity has minor effect on function of multiple synCAMs: Intβ1 and NCAM ICDs (linked to Fig. 1f).

(**a**) FACS analysis of αGFP synCAM expression in L929 fibroblast cells expressing cytosolic mCherry following cell sorting. The CAM TM and ICD domain for each construct is NCAM-1 or Intβ1, and the GFP-binding llama

nanobody (LaG) ECD for each construct is indicated. This analysis shows that this series of alternative affinity synCAMs are expressed at comparable levels. (**b**) Maximum projection of 20X confocal microscopy images of pairwise synCAM interfaces (t = 3 h, scale bar = 10 μm). Top: GFP-expressing cell (blue) is bound to an αGFP expressing cell (orange). The CAM TM and ICD domain for each pair is Intβ1. Bottom: GFP channel of the interfaces above highlighting differences of receptor enrichment at the interface. (**c**) Maximum projection of 20X confocal microscopy images of pairwise synCAM interfaces (t = 3 h, scale bar = 10 μm). Top: GFP-expressing cell (blue) is bound to an αGFP expressing cell (orange). The CAM TM and ICD domain for each pair is NCAM-1. Bottom: GFP channel of the interfaces above highlighting differences of receptor enrichment at the interface. (**d**) Plots of contact angles measured from the interfaces shown in b and c in relation to the corresponding LaG nanobody affinity (data are presented as mean values of of n = 10 pairs, error = 95 % CI). The contact angles for Intβ1 (blue) are shown in relation to NCAM-1 (red) and the tether control from Fig. 1f (black). (**e**) Plots of GFP enrichment measured from the interfaces shown in b and c in relation to the corresponding LaG nanobody affinity (data are presented as mean values of n = 10 pairs, error = 95 % CI). The GFP enrichment for NCAM-1 (red) are shown compared to Intβ1 (blue) and the tether control from Fig. 1f (black).

Source data

Extended Data Fig. 3 Differential sorting of synCAMs with varying ECD affinity and ICD (*linked to main* Fig. 1f).

(**a**) Cartoon depiction of the differential sorting competition assay (left) and quantification of radial distribution that is represented as a heat map (right). This experiment represents an alternative way to measure adhesion preferences/strength of the diverse synCAM-driven cell–cell interactions that differs from the contact angle measurement shown in Fig. 1f. Here we mix surface GFP L929 cells (bait cells) with two competing differentially labelled L929 cells, each with a different αGFP synCAM. Stronger adhesion of the synCAM is assessed via the relative degree of co-sorting of the competitor cells to the core in conjunction with the bait cells. We calculate the radial distribution of competing cells (red/blue) from the centroid of the

spheroid. (**b**) Representative maximum projection images of cell sorting competition assay between L929 cells expressing αGFP-ICAM-1 with the indicated ECD LaG nanobody (mCherry or BFP) mixed with L929 cells expressing GFP-ICAM-1 (t = 24 h, scale bar = 50 μm). (**c**) Quantification of the cell sorting competition assay from b (n = 4). (**d**) Representative maximum projection images of cell sorting competition assay between L929 cells expressing αGFP-ICAM-1 and cytosolic and L929 cells expressing αGFP-Tether mixed with L929 cells expressing GFP-ICAM-1 (scale bar = 20 μm, t = 24 h). (**e**) Quantification of the cell sorting competition assay from d (n = 4).

Source data

Extended Data Fig. 4 Characterization of tuning synCAM expression.

(**a**) FACS analysis of GFP synCAM and αGFP synCAM expression in L929 fibroblast cells following cell sorting with an ICAM-1 or Tether ICD. For the GFP constructs, expression is shown both for total GFP signal in the cell (Y-axis) and cells stained with an αFlag APC 647 antibody (x-axis). (**b**) Maximum projection of 20X confocal microscopy images of pairwise synCAM interfaces (t = 3 h, scale bar = 10 μm) of different expression levels from panel a: GFP-expressing cell (blue) is bound to an αGFP expressing cell (orange). The CAM TM and ICD domain for each pair is ICAM-1 or Tether. (**c**) Box and whisker plots (box = 25th to 75th percentile, whiskers = min to max, centre = median) of contact angles measured from the interfaces shown in b (n = 10 pairs).

Source data

Extended Data Fig. 5 Cell spreading with alternative synCAMs (linked to Fig. 2).

(a) Example microscopy images of cell spreading assays from Fig. 2, showing phenotypes for all synCAM species (Scale bar = 10 μm). Representative images are shown of independent replicates from Tether n = 10, ICAM-1 n = 20, JAM-B n = 20, MUC-4 n = 15, NCAM-1 n = 20, Intβ1 n

= 20, Intβ2 n = 20. SynCAMs are expressed in L929 fibroblasts and plated on a GFP coated glass surface. Cell footprint detected by membrane dye is indicated in blue outline; actin as stained by Phalloidin and shown in white. (b) Cartoon depicting the cell-spreading assay. L929 cells expressing an αGFP synCAM are plated on a GFP-coated surface and monitored over time. (c) Represented images from cell spreading assay of L929 cells expressing the indicated synCAMs. Individual slices from confocal images are shown. Scale bar = 10 µm. (d) Representative cell spreading contact area progress curves of L929 cells expressing the indicated synCAMs. Error = SEM. (e) Calculated spreading constants for L929 cells expressing the indicated synCAMs (where n is the number of unique cells analysed, Tether n = 24, Ecad n = 17, JAM-B n = 23, ICAM-1 n = 16, Intβ1 n = 16, Intβ2 n = 18, NCAM-1 n = 14, MUC-4 n = 12. Indicated line represents the median value).

Source data

Extended Data Fig. 6 How synCAM morphology is perturbed by small molecule inhibitors of specific actin regulators and loss of function mutations (linked to Fig. 2).

(a) Example microscopy images of L929 fibroblasts expressing αGFP JAM-B, ICAM-1, or Tether spreading on a GFP coated surface and stained with Phalloidin (scale bar = 10 µm). Spreading is shown in the presence of the indicated inhibitor of actin regulation. A minimum of 10 regions of interest were imaged on two separate days. (b) Maximum projection confocal images (scale bar = 10 µm), and calculated contact angles of synCAM interfaces containing the ICAM-1 ICD with mutations in the ERM binding domains (BD) (box = 25th to 75th percentile, whiskers = min to max, centre = median, n = 20 pairs)[60]. (c) Maximum projection confocal images (scale bar = 10 µm), and calculated contact angles of synCAM interfaces containing the Intβ1 ICD with mutations in the two “NPxY” talin binding domain motifs (box = 25th to 75th percentile, whiskers = min to max, centre = median, n = 20 pairs)[61]. (d) Maximum projection confocal images (scale bar = 10 µm), and calculated contact angles of synCAM interfaces containing the Intβ2 ICD with mutations in the two “NPxF” talin binding domain motifs (box = 25th to 75th percentile, whiskers = min to max, centre

= median, n = 20 pairs)[62]. (e) Maximum projection confocal images (scale bar = 10 µm), and calculated contact angles of synCAM interfaces containing the Ecad ICD with mutations in the β -catenin binding domain (box = 25th to 75th percentile, whiskers = min to max, centre = median, n = 20 pairs)[63]. (f) Maximum projection confocal images (scale bar = 10 µm), and calculated contact angles and GFP enrichment of synCAM interfaces containing the MUC-4 ICD with mutations in Ser and Tyr phosphorylation sites (box = 25th to 75th percentile, whiskers = min to max, centre = median, n = 20 pairs). (g) Maximum projection confocal images (scale bar = 10 µm), and calculated contact angles and GFP enrichment of synCAM interfaces containing the JAM ICD with mutations in the PDZ binding domain (box = 25th to 75th percentile, whiskers = min to max, centre = median, n = 20 pairs). (h) Maximum projection confocal images (scale bar = 10 µm), and calculated contact angles and GFP enrichment of synCAM interfaces containing the NCAM-1 ICD with mutations in the Cys palmitoylation site (box = 25th to 75th percentile, whiskers = min to max, centre = median, n = 20 pairs)[64].

Source data

Extended Data Fig. 7 Asymmetric cell–cell interfaces: mismatched ICDs (*linked to main* Fig. 3).

(**a**) Maximum projection of 20x confocal microscopy images of pairwise synCAM interfaces (scale bar = 10 µm, t = 3 h): GFP-expressing cell (blue) is bound to an αGFP expressing cell (orange) containing the indicated CAM ICD. Representative images are shown of 10 independent cell pairs. (**b**) GFP channel of cell pairs shown in a.

Extended Data Fig. 8 Testing orthogonality of synCAM ECD pairs by sorting assays (*linked to main* Fig. 4a).

(**a**) Cartoon depicting differential sorting assay used to determine orthogonality of synCAM ECD pairs. SynCAM pairs are mixed with parental L929 cells and imaged after 24 h. Sorting of parental cells should only occur if the cognate synCAM ECDs are correctly matched and able to bind. (**b**) Representative maximum projection images of differential sorting

assay for a subset of the synCAMs with orthogonal ECDs (scale bar = 20 µm). Parental L929 cell sorting was only observed in the case of matching ECDs. (**c**) Quantification of sorting from **b** (n = 6). The difference of average distance from the centre of the sphere between parental L929 cells and BFP+ cells were calculated and are represented as a heat map. Exclusion of parental cells is observed in the case of matching synCAM pairs. (**d**) Representative maximum projection images synCAM design containing multiple epitopes within a single ECD (scale bar = 20 µm). The HA-CD19 ECD exhibits differential sorting for either αCD19 or αHA synCAMs only. Thus, we can generate OR-gate synCAMs capable of pairing with multiple adhesion partners. (**e**) Quantification of sorting from **d** (n = 6). The difference of average distance from the centre of the sphere between parental L929 cells and BFP+ cells were calculated and are represented as a heat map. Exclusion of parental cells is observed in the case of matching synCAM pairs.

Source data

Extended Data Fig. 9 Replicates and distribution of assemblies formed from custom homotypic synCAMs (linked to Fig. 4d).

Maximum projection of 20X confocal microscopy images of differential sorting between L929 cells expressing WT Ecad or the indicated homophilic-binding synCAMs (scale bar = 50 µm, t = 48 h). Representative images, assembly classifications and distributions are shown for Ecad-IF1 (a), Ecad-Aph4 (b), IF1-Aph4 (c) and Ecad-IF1-Aph4 (d).

Extended Data Fig. 10 Targeting WT Pcad with synCAM (*linked to main* Fig. 4e).

(a) Maximum projection images of the sorting assay in which L929 cells expressing WT Pcad (orange) are mixed with parental (left) or synCAM (right) L929 cells (blue, t = 0, 24 h, scale bar = 50 µm). (b) Quantification of the relative area between the BFP negative control or αPcad synCAM and mCherry (Pcad) L929 cells over the course of the 24-h assembly (n = 4 biologically independent samples, error = SEM). A greater difference in area is consistent with a more compact Pcad sphere and exclusion of BFP+ cells.

(c) Quantification of relative distance per cell (BFP-mCherry) from the centre of the sphere following assembly (t = 24 h, n = 4 biologically independent samples, line= mean). WT BFP cells exhibit a greater difference in distance, which is consistent with their exclusion from the Pcad sphere, while αPCAD synCAMs intercalate.

Source data

Extended Data Fig. 11 SynCAMs function in primary cells: mesenchymal stem cells (MSCs), dermal fibroblasts, and iPSC derived smooth muscle cells (*linked to main* Fig. 4).

Maximum projection of 20x confocal microscopy images of αGFP and GFP synCAMs (with ICAM-1 ICD) or corresponding tether (no ICD) expressed in MSCs (a, scale bar = 10 µm) primary dermal fibroblasts (b, scale bar = 20 µm) or iPSC derived SMCs (c, scale bar = 20 µm). αGFP cells were also labelled with mCherry; GFP cells were also labelled with BFP. Representative images are shown of three independent replicates. In both cell types, the GFP-tether is diffusely spread throughout the cell. In contrast, the GFP-synCAM is strongly enriched atheterotypic cell–cell interfaces (white arrows). When cells expressing GFP-synCAMs are plated without their partner cells, the GFP is diffusely distributed throughout the cell.

Extended Data Fig. 12 Control of Multicellular organization by synCAMs (*linked to main* Fig. 5).

(a) cartoon depicting modulation of WT Ncad (green) and WT Pcad (orange) sorting through introduction of synCAMs. (b) Maximum projections of 20X confocal microscopy images of WT Pcad and WT Ncad L929 cells with expression of the indicated heterophilic synCAMs (scale bar = 20 µm, t = 24 h). The GFP-synCAM is expressed in the Ncad-expressing L929 cell and αGFP synCAM in the Pcad-expressing L929 cell. Representative images are shown of three independent replicates. This data shows that synCAMs can drive integration between differentially sorting Pcad and Ncad cells, just as they can between Pcad and Ecad cells (Fig. 5a). (c) Quantification of roundness (left) and total surface area (right) of L929 cells from maximum projections of 20x confocal images in Fig. 5b (data are

presented as mean values of n = 18 unique fields analysed across two independent wells, error = SD). (d) 3D (top) and maximum projection (bottom) views of multicellular assemblies between a mouse intestinal epithelial monolayer (green) and mouse embryonic fibroblast cells (MEFs) (orange) with either a GFP-αGFP tether (left) or synthetic ICAM-1 (right) heterophilic adhesion interaction. Representative images are shown of two independent replicates.

Source data

Supplementary information

Reporting Summary

Supplementary Table 1

SynCAM cell–cell interfaces. Contact angle and GFP enrichment measurements are provided. Error = 95% CI.

Supplementary Table 2

Sequences of protein constructs used in this study. The epitope tag is highlighted in red, the ECD-binding region is shown in blue, and the TM region and ICD in bold.

Supplementary Table 3

SLiM domains from CAM ICDs used in synCAMs. SLiMs were identified from ELM database and literature sources.

Supplementary Video 1

Self-assembly of cells with orthogonal synCAM pairs. Alternative ECD with ICAM-1 ICD are shown (linked to Fig. 4a).

Supplementary Video 2

Three-cell synCAM interaction network. Three-cell synCAM interaction networks drive cell assembly (linked to Fig. 4b).

Supplementary Video 3

SynCAM intercalation into native assemblies. Anti-PCAD synCAM (ICAM-1 ICD) drives cell intercalation into PCAD cluster (linked to Fig. 4e).

Supplementary Video 4

Remodelling tissue organization. SynCAMs can drive coupling and complex remodelling of epithelial and spheroid tissues (linked to Fig. 5b).

Source data

Source Data Fig. 1

Source Data Fig. 2

Source Data Fig. 3

Source Data Fig. 4

Source Data Extended Data Fig. 2

Source Data Extended Data Fig. 3

Source Data Extended Data Fig. 4

Source Data Extended Data Fig. 5

Source Data Extended Data Fig. 6

Source Data Extended Data Fig. 8

[Source Data Extended Data Fig. 10]

[Source Data Extended Data Fig. 12]

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Stevens, A.J., Harris, A.R., Gerdts, J. *et al.* Programming multicellular assembly with synthetic cell adhesion molecules. *Nature* **614**, 144–152 (2023). https://doi.org/10.1038/s41586-022-05622-z

- Received: 28 October 2021
- Accepted: 02 December 2022
- Published: 12 December 2022
- Issue Date: 02 February 2023

- DOI: https://doi.org/10.1038/s41586-022-05622-z

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

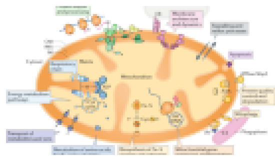
Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

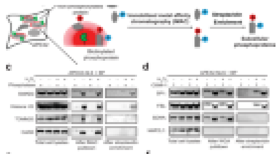
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05622-z



Mitochondrial proteins: from biogenesis to functional networks

09 January 2019

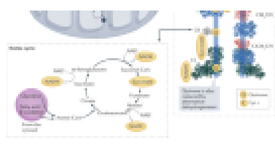
Nikolaus Pfanner, Bettina Warscheid & Nils Wiedemann



Spatiotemporally-resolved mapping of RNA binding proteins via functional proximity labeling reveals a mitochondrial mRNA anchor promoting stress recovery

17 August 2021

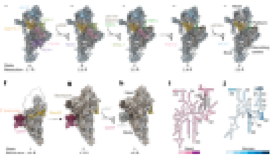
Wei Qin, Samuel A. Myers, … Alice Y. Ting



The assembly, regulation and function of the mitochondrial respiratory chain

07 October 2021

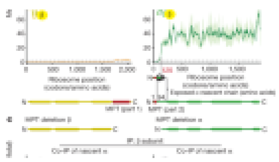
Irene Vercellino & Leonid A. Sazanov



[Principles of mitoribosomal small subunit assembly in eukaryotes]

08 December 2022

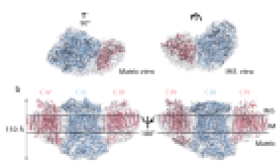
Nathan J. Harper, Chloe Burnside & Sebastian Klinge



[Cotranslational assembly of protein complexes in eukaryotes revealed by ribosome profiling]

29 August 2018

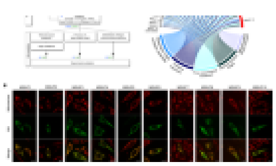
Ayala Shiber, Kristina Döring, … Bernd Bukau



[Cryo-EM structure of the yeast respiratory supercomplex]

31 December 2018

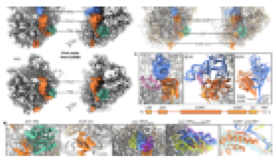
Sorbhi Rathore, Jens Berndtsson, … Martin Ott



[Mitochondrial peptide BRAWNIN is essential for vertebrate respiratory complex III assembly]

11 March 2020

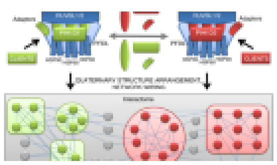
Shan Zhang, Boris Reljić, … Lena Ho



[Structure and function of Vms1 and Arb1 in RQC and mitochondrial proteome homeostasis]

12 June 2019

Ting Su, Toshiaki Izawa, … Roland Beckmann



[How do our cells build their protein interactome?]

27 July 2018

Benoit Coulombe, Philippe Cloutier & Marie-Soleil Gauthier

- Article
- Open Access
- Published: 25 January 2023

Mitochondrial complexome reveals quality-control pathways of protein

import

- Uwe Schulte,
- Fabian den Brave,
- Alexander Haupt,
- Arushi Gupta,
- Jiyao Song,
- Catrin S. Müller,
- Jeannine Engelke,
- Swadha Mishra,
- Christoph Mårtensson,
- Lars Ellenrieder,
- Chantal Priesnitz,
- Sebastian P. Straub,
- Kim Nguyen Doan,
- Bogusz Kulawiak,
- Wolfgang Bildl,
- Heike Rampelt,
- Nils Wiedemann,
- Nikolaus Pfanner ,
- Bernd Fakler &
- …
- Thomas Becker

Show authors

Nature **volume 614**, pages 153–159 (2023)

- 16k Accesses

- 176 Altmetric

- Metrics details

Abstract

Mitochondria have crucial roles in cellular energetics, metabolism, signalling and quality control[1,2,3,4]. They contain around 1,000 different proteins that often assemble into complexes and supercomplexes such as respiratory complexes and preprotein translocases[1,3,4,5,6,7]. The composition of the mitochondrial proteome has been characterized[1,3,5,6]; however, the organization of mitochondrial proteins into stable and dynamic assemblies is poorly understood for major parts of the proteome[1,4,7]. Here we report quantitative mapping of mitochondrial protein assemblies using high-resolution complexome profiling of more than 90% of the yeast mitochondrial proteome, termed MitCOM. An analysis of the MitCOM dataset resolves >5,200 protein peaks with an average of six peaks per protein and demonstrates a notable complexity of mitochondrial protein assemblies with distinct appearance for respiration, metabolism, biogenesis, dynamics, regulation and redox processes. We detect interactors of the mitochondrial receptor for cytosolic ribosomes, of prohibitin scaffolds and of respiratory complexes. The identification of quality-control factors operating at the mitochondrial protein entry gate reveals pathways for preprotein ubiquitylation, deubiquitylation and degradation. Interactions between the peptidyl-tRNA hydrolase Pth2 and the entry gate led to the elucidation of a constitutive pathway for the removal of preproteins. The MitCOM dataset—which is accessible through an interactive profile viewer—is a comprehensive resource for the identification, organization and interaction of mitochondrial machineries and pathways.

Main

Mitochondria are multifunctional organelles. In addition to their roles in oxidative phosphorylation and in the metabolic pathways of amino acids, lipids, haem and iron–sulfur clusters, they perform functions in cellular signalling, redox processes, quality control and apoptosis[1,2,3]. Most mitochondrial proteins are imported as precursors from the cytosol, whereas about 1% of the proteins are synthesized inside the organelle. Mitochondria are dynamic organelles that frequently divide and fuse, and they display a characteristic folded structure of their inner membrane. Defects in the mitochondria can lead to severe diseases, particularly of the central nervous system, metabolism and the cardiovascular system[2,3].

The protein complement of the mitochondria has been determined in systematic proteomic studies[3,4,5], with a coverage of more than 90% for the mitochondrial proteome in the model organism baker's yeast (*Saccharomyces cerevisiae*)[6]. By contrast, the organization of mitochondrial proteins into protein assemblies, from stable complexes and supercomplexes to transient assembly intermediates, is only partially understood. Various approaches such as affinity purification, native electrophoresis, gel filtration, density gradients, cross-linking and structural biology have been applied to study the organization of the mitochondrial proteome[1,4,7]. Each of these approaches provided important information on selected mitochondrial complexes, but none yielded a comprehensive overview of the expected large number of distinct protein assemblies.

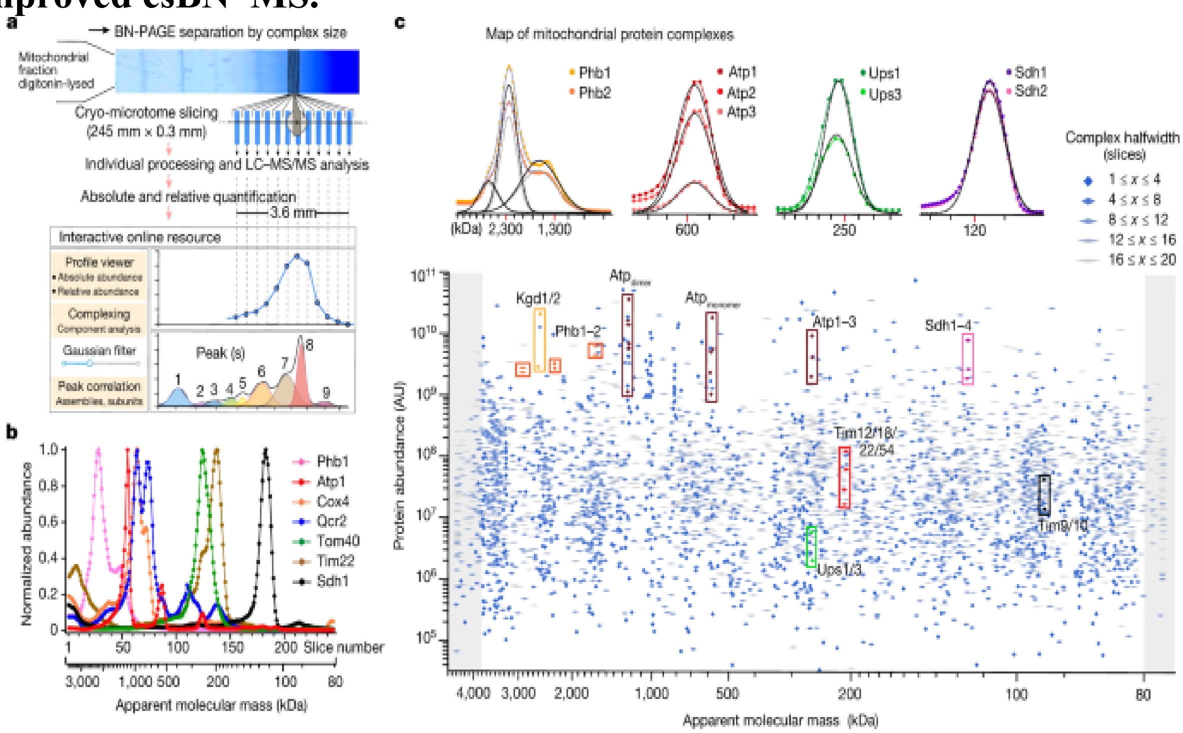
We report a comprehensive high-resolution complexome of yeast mitochondria and associated proteins, based on blue native electrophoresis combined with cryo-slicing and mass spectrometry analysis (csBN–MS)[8,9]. We systematically improved csBN–MS from protein separation to advanced detection and quantification of protein profiles, yielding the high resolution and coverage of the mitochondrial complexome (MitCOM)[7,9,10]. MitCOM covers more than 90% of the high-confidence mitochondrial proteome and provides a wealth of information on mitochondrial protein assemblies and their quantitative appearance.

High-resolution mitochondrial complexome

Yeast mitochondria were lysed with the non-ionic detergent digitonin and processed for blue native gel electrophoresis, high-resolution cryo-slicing and quantitative MS (Fig. 1a and Extended Data Figs. 1 and 2). We optimized the conditions for protein separation, quantification and evaluation to substantially improve the resolution and proteome coverage, including (1) preparation and lysis of mitochondria under mild conditions, followed by separation of protein assemblies by blue native electrophoresis; (2) cryo-slicing of blue native gel lanes into 245 slices with a width (0.3 mm) three times smaller than the half-width of the sharpest-focusing protein peaks; and (3) elaborate MS quantification and profile building (Extended Data Fig. 1 and Methods). We determined abundance–mass profiles for 1,891 different proteins, 906 of them represented high-confidence

mitochondrial proteins and made up around 96% of the protein mass in our preparation[6] (MitCOM; Supplementary Tables 1 and 2), whereas another 985 profiles originated from non-mitochondrial proteins representing about 4% of the total protein mass (Extended Data Fig. 2a and Supplementary Table 3). MitCOM covers a molecular mass range from around 80 kDa to 3,800 kDa with homogenous resolution and a protein abundance range of more than six orders of magnitude (Fig. 1b,c, Extended Data Fig. 2c,d and Supplementary Tables 1 and 2). The individual abundance–mass profile of each MitCOM protein represents a protein-specific fingerprint with a characteristic shape for the distribution and abundance of this protein in one or more assemblies (peaks) (Fig. 1b and Extended Data Fig. 2b,e).

Fig. 1: High-resolution complexome profiling of mitochondria using improved csBN–MS.



a, Schematic of the csBN–MS workflow, comprising protein separation by blue native gel electrophoresis (BN-PAGE), cryo-slicing of the gel, MS analysis, data processing and protein quantification. The interactive online platform presents all resulting abundance–mass profiles and MS data for user-directed inspection and evaluation. LC–MS/MS, liquid chromatography coupled with tandem MS. **b**, Abundance profiles over slice number and apparent molecular mass for the indicated set of mitochondrial proteins; the

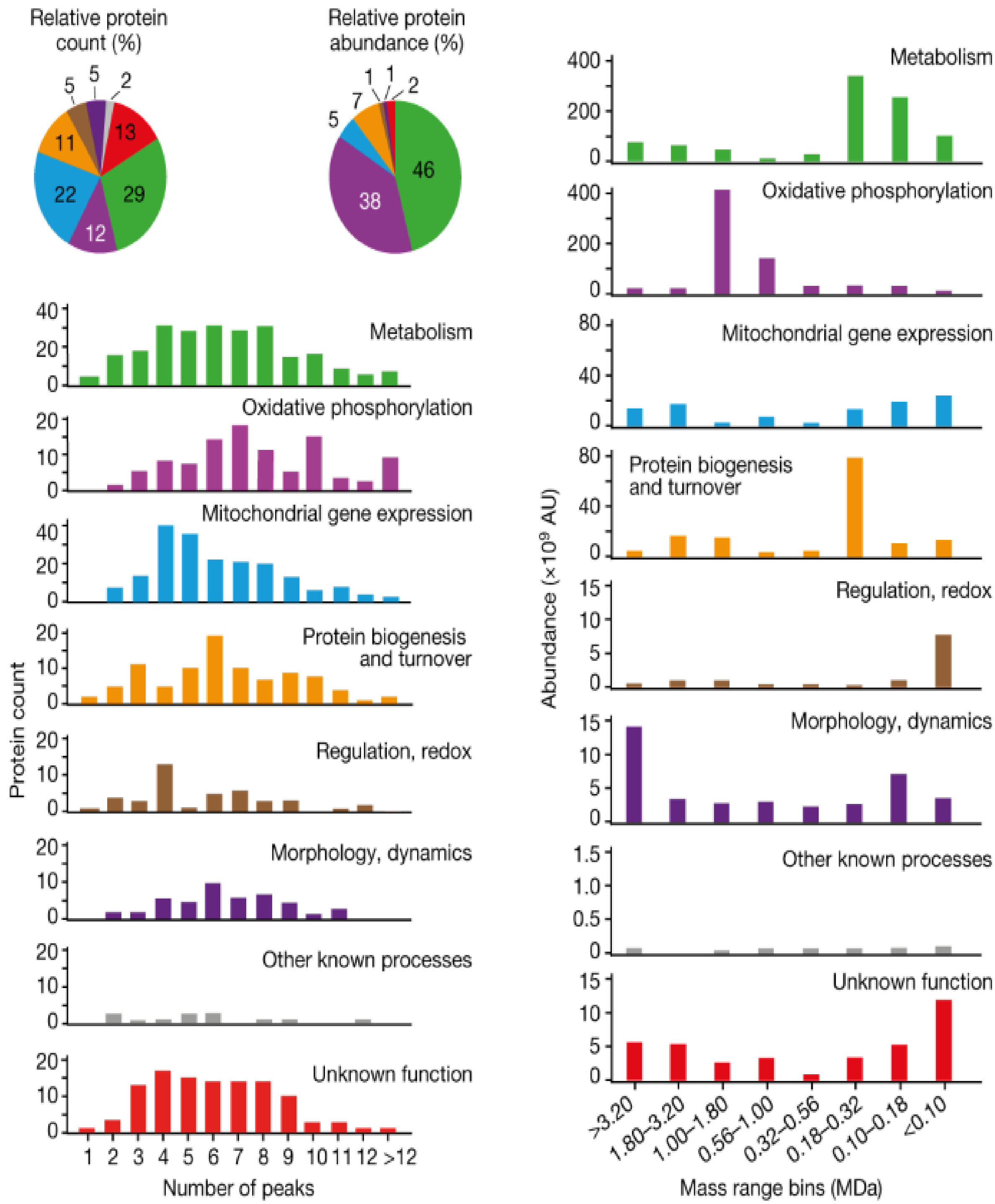
profiles were normalized to the maximal abundance value determined for the respective protein. **c**, Map of protein peaks determined from abundance–mass profiles by multi-Gaussian fits and verified by manual inspection. All of the symbols represent identified peak parameters (apparent molecular mass, abundance and half-width as indicated at the top right). The shaded areas denote peaks with maxima outside the blue native gel. Coloured frames denote known protein complexes that are assembled from the indicated constituents of which the peaks are highlighted in the same colour. Insets: sections of abundance–mass profiles determined for the indicated protein constituents of established mitochondrial assemblies together with the fit lines (dashed) obtained by Gaussian fitting. For Phb1 and Phb2, the composite fit lines (solid, blue) and the individual components (dashed, grey, black) are shown. Note the close match in the midpoint of the individual Gaussian functions obtained in independent fits for the constituents of the same complex.

We performed an unbiased automated component analysis based on peak detection and the fitting of multi-Gaussian functions to the abundance–mass profiles, followed by rigorous quality control by manual inspection. We identified a combined 5,224 peaks in the profiles of the 818 MitCOM proteins accessible to the fitting procedure (Fig. 1c and Extended Data Fig. 2e). The number of identified protein peaks substantially exceeds previous complexome profiling efforts in mitochondria[7,9,11]. The MitCOM dataset is presented in an openly accessible platform for interactive analysis through the integrated profile viewer (https://www.complexomics.org/datasets/mitcom or https://www3.cmbi.umcn.nl/cedar/browse/experiments/CRX36; Extended Data Fig. 3). The resolution and accuracy of the MitCOM profiles promoted unsupervised systemic analysis of assemblies and their composition, a major challenge in the exploration of native protein complexes[11,12]. A distance measure combining information on peak intensity, mass range and correlation coefficients (Extended Data Fig. 4 and Methods) was used for the automated discrimination of protein complex components by *t*-distributed stochastic neighbourhood distribution (*t*-SNE) (Extended Data Figs. 4 and 5).

Complexity of protein organization

An analysis of the MitCOM dataset showed that the majority of mitochondrial proteins appeared as constituents of several assemblies with an average of 6.4 peaks per protein (complexity), whereas only about 1% of the proteins exhibited a single peak (Fig. 2 and Extended Data Fig. 6a). The complexity was largely independent of protein properties such as the number of predicted transmembrane segments and the molecular mass, and only moderately correlated with protein abundance (Extended Data Fig. 6b). However, complexity as well as molecular mass ranges of protein assemblies markedly differed between the four mitochondrial subcompartments—the outer membrane, intermembrane space, inner membrane and matrix (Extended Data Fig. 6c). The functional classification of MitCOM protein assemblies revealed characteristic differences (Fig. 2). The highest mean number of peaks per protein was observed for the subcategories cofactor biosynthesis (for example, coenzyme Q biosynthesis[4,6]) and oxidative phosphorylation assembly (Extended Data Fig. 6d). The molecular mass distribution of protein assemblies differed substantially between the functional categories such as metabolism, oxidative phosphorylation, protein biogenesis, regulatory processes and morphology (Fig. 2). Proteins with an as yet unknown function, representing only around 2% of the MitCOM mass but around 13% of the different MitCOM proteins, were found in low- and high-molecular-mass ranges. Thus, small as well as very large mitochondrial protein assemblies contain a considerable number of proteins that lack functional annotation.

Fig. 2: Complexity and size of mitochondrial protein complexes related to functional categories.



Summary of the complexity (left, number of peaks per protein) or the apparent molecular size (right, mass range bins) of assemblies/complexes identified in the abundance–mass profiles of 815 proteins assigned to distinct categories of mitochondrial function[6] (Extended Data Fig. 6). The relative numbers (left) and abundances (right) of the functional categories

are shown. The apparent mass range was binned into equal log-value intervals (bins of 0.25) of the apparent molecular mass.

The non-mitochondrial proteins in the source preparation showed an average complexity of 4.5 peaks per protein and predominantly originated from cellular compartments in the vicinity of mitochondria, cytosol, endoplasmic reticulum and nucleus (Extended Data Figs. 2a and 7 and Supplementary Table 3). Protein biogenesis, degradation and quality control represented the largest functional category, consistent with the identification of quality-control factors associated with mitochondria outlined below.

Taken together, the MitCOM dataset uncovered a notable complexity of mitochondrial protein assemblies over an extended molecular mass and abundance range, suggesting the occurrence of many as yet unknown protein complexes and interactions of this organelle.

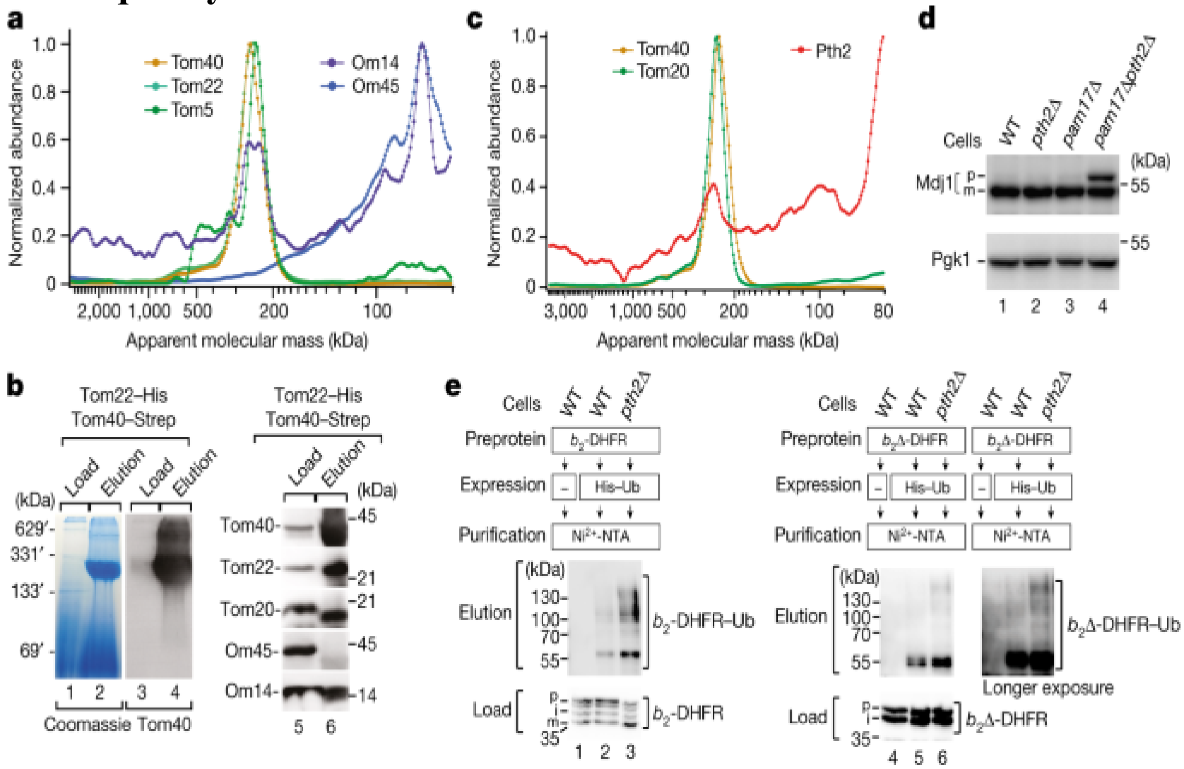
Dynamic protein assemblies

MitCOM demonstrated high precision in the analysis of dynamic protein machineries such as the outer membrane sorting and assembly machinery (SAM), and the separation of complexes of different abundance, such as the carrier translocase of the inner membrane (TIM22) and succinate dehydrogenase (SDH, respiratory complex II) (Extended Data Fig. 8a–c). Similarly, MitCOM quantitatively separated distinct complexes and subcomplexes of the F_1F_O-ATP synthase (Extended Data Fig. 8d).

We examined whether MitCOM can be used to identify unknown binding partners of protein assemblies. A population of the outer membrane protein Om14 was found to co-migrate with the translocase of the outer membrane (TOM) complex at around 250 kDa, whereas another population of Om14 migrated together with its established interaction partner Om45 in a low-molecular-mass assembly[13,14] (Fig. 3a). We verified the interaction between Om14 and TOM using two-step affinity purification, which yielded a highly purified TOM complex containing Om14 but not Om45 (Fig. 3b). Om14 was reported to function as a receptor for translating cytosolic ribosomes[15]; however, it has been unclear how Om14-bound ribosomes can transfer nascent precursor polypeptides to the mitochondrial entry gate TOM.

MitCOM shows that Om14 is present in two populations, Om14–Om45 and Om14–TOM, revealing the missing direct link of Om14 to the protein import site. Thus, the ribosome receptor can cooperate with TOM in co-translational protein import.

Fig. 3: The TOM complex interacts with the ribosome receptor Om14 and the quality-control factor Pth2.



a, Normalized abundance–mass profiles of the TOM subunits Om14 and Om45. **b**, Tom22–His Tom40–Strep mitochondria were lysed with digitonin and processed for tandem affinity purification through Ni-NTA agarose and Strep-Tactin Sepharose. The samples were analysed by blue native electrophoresis or SDS–PAGE followed by Coomassie staining (lanes 1 and 2) or immunodetection using the indicated antisera. Load, 0.2% (blue native gel) or 1% (SDS–PAGE); elution, 100%. **c**, Normalized abundance–mass profiles of Tom40, Tom20 and Pth2. **d**, Cell extracts of the indicated strains were analysed by immunodetection. p, precursor; m, mature form. **e**, Wild-type (WT) and *pth2*Δ cells expressing cytochrome b_2–DHFR or b_2Δ–DHFR precursors and His-tagged ubiquitin as indicated were lysed under denaturing conditions and affinity-purified through Ni-NTA agarose. Proteins were analysed by SDS–PAGE and immunodetection. Affinity

purification of tagged ubiquitin under denaturing conditions leads to an enrichment of proteins with covalently attached ubiquitin[41]. b_2-DHFR–Ub, ubiquitin modified b_2-DHFR; $b_2\Delta$-DHFR–Ub, ubiquitin modified $b_2\Delta$-DHFR; i, intermediate.

To study whether MitCOM can reveal new interaction partners of very large assemblies, we analysed inner membrane complexes. We identified the yeast J protein Mdj2 as a partner of the prohibitin–*m*-AAA supercomplex of about 2.2 MDa (Extended Data Fig. 8e,f). By comparing the MitCOM profiles of assembly factors of cytochrome *c* oxidase (COX, respiratory complex IV), we found that a main peak of Shy1 is present at the fully assembled III_2IV_2 respiratory supercomplex of 1 MDa, different from other assembly factors such as Mss51 and Coa1 (Extended Data Fig. 8g,h). We conclude that the high resolution and quantitative nature of MitCOM make it a powerful source for the analysis of dynamic protein machineries and the identification of interaction partners.

Pth2 is involved in the removal of preproteins

We investigated whether MitCOM can be used to identify pathways that involve dynamic interactions and components of low abundance, such as quality-control pathways that monitor mitochondrial proteostasis and remove accumulated preproteins[16,17,18,19,20] (Extended Data Fig. 9a). We observed a co-migration of the peptidyl-tRNA hydrolase Pth2 with the TOM complex (Fig. 3c). Pth2 is a mitochondrial outer-membrane protein[6,21]. In addition to its peptidyl-tRNA hydrolase activity[22,23], Pth2 interacts with cytosolic factors of the ubiquitin proteasome system[24]. It has been unknown whether mature Pth2 has mitochondrial interaction partners and whether it has a role in mitochondrial function or quality control.

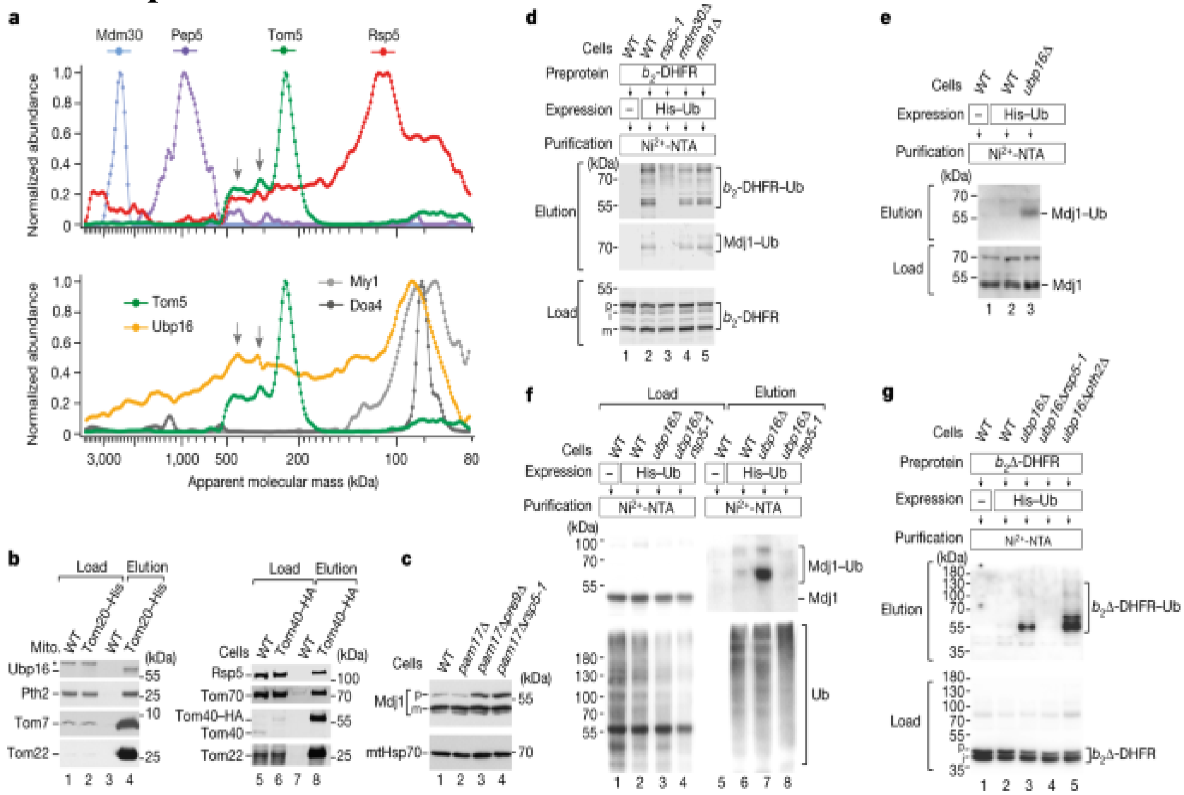
Affinity purification through tagged Tom40 as well as tagged Pth2 demonstrated the association between Pth2 and the TOM complex (Extended Data Fig. 9b,c). Deletion of the *PTH2* gene caused a growth defect on non-fermentable medium on which mitochondrial respiration is critical for cell survival, and a double-mutant strain *pth2*Δ*pam17*Δ displayed a synthetic growth defect (Extended Data Fig. 9d). Pam17 is a subunit of the

presequence translocase-associated motor (PAM) of the mitochondrial protein import machinery[25,26]. Loss of Pam17 reduces the protein import efficiency and causes an accumulation of mitochondrial preproteins when the degradation of non-imported proteins is inhibited[17]. The precursor form of Mdj1 accumulated in *pth2*Δ*pam17*Δ cells (Fig. 3d), indicating that the removal of non-imported preproteins was impaired. To analyse whether accumulated preproteins were ubiquitylated, we used a yeast strain expressing tagged ubiquitin and cytochrome b_2 precursor variants containing dihydrofolate reductase (DHFR) (Methods). In *pth2*Δ-mutant cells, increased amounts of ubiquitylated preproteins accumulated (Fig. 3e). Thus, the lack of Pth2 leads to an accumulation and ubiquitylation of mitochondrial preproteins, suggesting that Pth2 is involved in the removal of preproteins from the import-site TOM.

Precursor ubiquitylation at the TOM complex

The accumulation of ubiquitylated preproteins at the mitochondrial entry gate[17,27] (Fig. 3e) and the ubiquitylation-mediated modulation of mitochondrial protein biogenesis in mammalian cells[27,28] suggest that ubiquitylation and deubiquitylation are important in the removal of preproteins from TOM. However, neither an E3 ubiquitin ligase nor a deubiquitylase operating at the yeast TOM complex have been identified. We extended the target-focused screening of MitCOM to proteins that are potentially involved in ubiquitylation and deubiquitylation. Out of several candidate proteins, populations of the cytosolic HECT-type E3 ubiquitin ligase Rsp5 (‘reverses *spt*⁻ phenotype’) and of the outer-membrane-localized ubiquitin-specific protease 16 (Ubp16) co-migrated with the TOM complex (Fig. 4a). Affinity purification with tagged TOM subunits confirmed the interactions of Rsp5 and Ubp16 with TOM (Fig. 4b). Several functions have been assigned to Rsp5, including regulation of biosynthesis of unsaturated fatty acids, ubiquitylation of endoplasmic reticulum–mitochondria contact sites, and the fusion GTPase Fzo1, protein degradation in the cytosol and multivesicular body formation[29,30,31,32]. Ubp16 is a predicted, yet lacking characterization, yeast deubiquitylase (homologue of USP30) with an amino-terminal anchor in the outer mitochondrial membrane[21,33].

Fig. 4: Rsp5 and Ubp16 control ubiquitylation of mitochondrial precursor proteins.



a, Normalized abundance–mass profiles of the TOM subunit Tom5, the E3 ubiquitin ligases Mdm30, Pep5 and Rsp5 (top), and the deubiquitylating enzymes Ubp16, Doa4 and Miy1 (bottom). The arrows depict profile peaks of co-migration of Tom5, Rsp5 and Upb16. **b**, WT and Tom20–His mitochondria (left) or WT and Tom40–HA cell extracts (right) were lysed with digitonin and affinity-purified through Ni-NTA agarose or anti-HA affinity matrix. Proteins were analysed using SDS–PAGE and immunodetection. Load, 0.5%; elution, 100%. The asterisk marks an unspecific signal of anti-Ubp16. **c**, Cell extracts of the indicated strains were analysed by immunodetection. p, precursor; m, mature form of Mdj1. Pre9, proteasomal subunit. **d**, WT, *rsp5-1*, *mdm30*Δ and *mfb1*Δ cells expressing cytochrome b_2-DHFR and His-tagged ubiquitin as indicated were lysed under denaturing conditions and affinity-purified through Ni-NTA agarose. Proteins were analysed using SDS–PAGE and immunodetection. Load, 0.2%; elution, 100%. **e**,**f**, WT and *ubp16*Δ (**e**) and WT, *ubp16*Δ and *ubp16*Δ *rsp5-1* (**f**) strains expressing His-tagged ubiquitin were lysed under denaturing conditions and affinity-purified through Ni-NTA agarose.

Proteins were analysed using SDS–PAGE and immunodetection. Load, 0.2%; elution, 100%. **g**, WT, *ubp16*Δ, *ubp16*Δ *rsp5-1* and *ubp16*Δ*pth2*Δ cells expressing cytochrome b_2Δ-DHFR and His-tagged ubiquitin as indicated were lysed under denaturing conditions and affinity-purified through Ni-NTA. Proteins were analysed using SDS–PAGE and immunodetection. Load, 0.2%; elution, 100%.

We analysed whether Rsp5 and Ubp16 are involved in the removal of mitochondrial preproteins. Deletion of *PAM17* in the *rsp5-1* yeast mutant resulted in a synthetic growth defect and accumulation of the Mdj1 precursor (Fig. 4c and Extended Data Fig. 9e), indicating that Rsp5 has a role in the removal of non-imported preproteins. Using yeast with tagged ubiquitin, we observed ubiquitylated b_2-DHFR and Mdj1 precursors. Ubiquitylation of both precursors was blocked in *rsp5-1* cells (Fig. 4d). As control, loss of Mdm30 and Mfb1, which are F-box protein subunits of SCF ubiquitin ligase complexes that regulate mitochondrial morphology[34,35], did not impair the ubiquitylation of the precursors (Fig. 4d). Thus, Rsp5 is required for the ubiquitylation of mitochondrial preproteins at the TOM complex (Extended Data Fig. 9f–h).

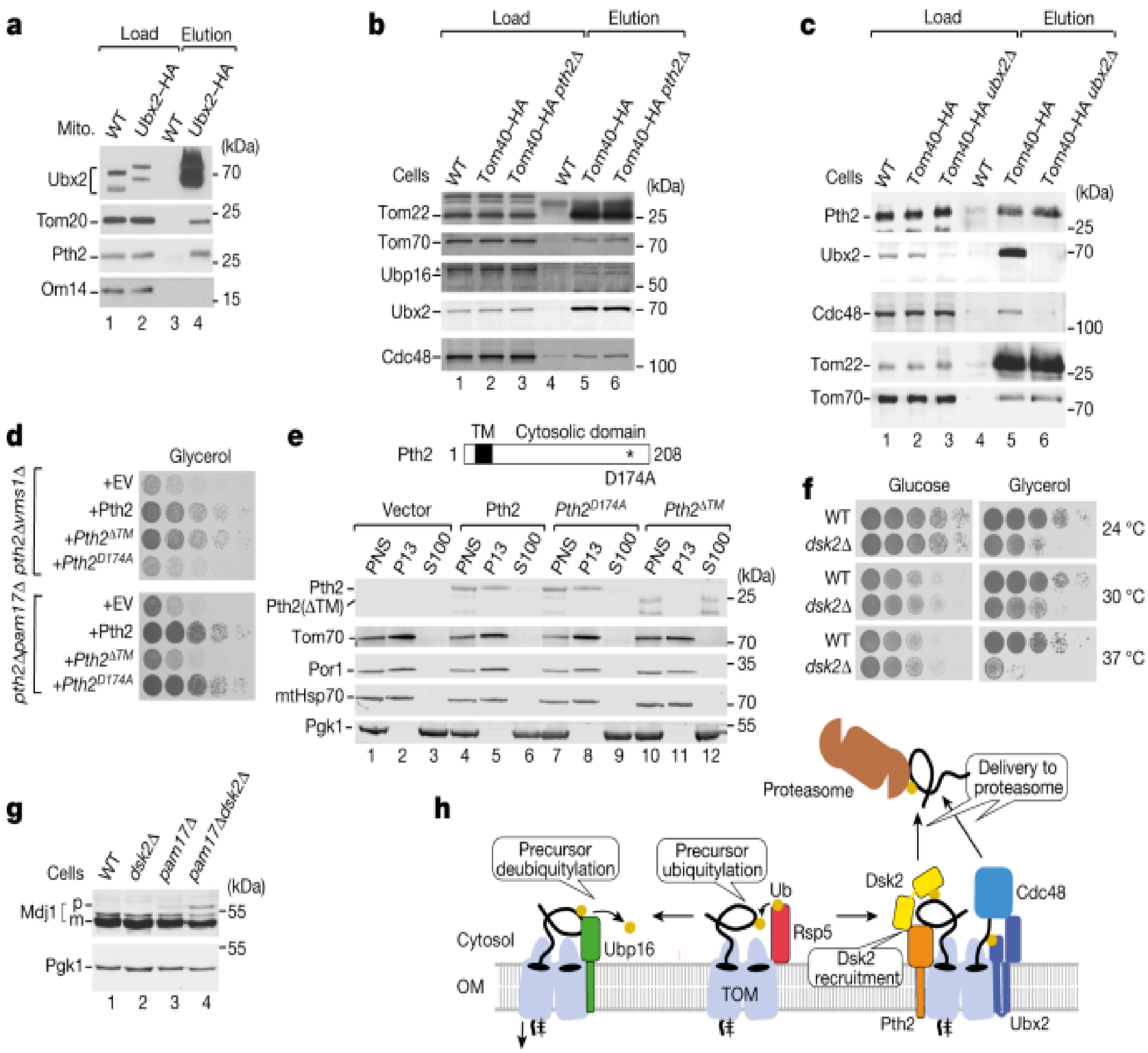
In mutant cells lacking the predicted deubiquitylase Ubp16, ubiquitylated forms of the Mdj1 and b_2Δ-DHFR precursors accumulated (Fig. 4e–g and Extended Data Fig. 10a,b). By combining *ubp16*Δ with the *rsp5-1* mutant, the ubiquitylation of Mdj1 and b_2Δ-DHFR was blocked (Fig. 4f,g), demonstrating that ubiquitylation by Rsp5 was a prerequisite for subsequent deubiquitylation of accumulated preproteins by Ubp16. The E3 ubiquitin ligases Doa10 and Hrd1 were, like Mdm30 and Mfb1, dispensable for the ubiquitylation of accumulated b_2Δ-DHFR (Extended Data Fig. 10c).

We conclude that Rsp5 and Ubp16 function in the ubiquitylation and deubiquitylation of mitochondrial preproteins at the TOM complex. In the *pth2*Δ*ubp16*Δ double mutant, the ubiquitylation of b_2Δ-DHFR was enhanced further (Fig. 4g). Together with the accumulation of ubiquitylated preproteins in *pth2*Δ-mutant cells (Fig. 3e), this indicates that Pth2 is involved in the removal of ubiquitylated precursors.

Pth2 and mitochondrial quality control

We examined whether Pth2 is part of an already known quality-control pathway of mitochondria or whether it constitutes a separate pathway for the removal of preproteins. First, the mitochondrial protein translocation-associated degradation (mitoTAD) pathway removes non-imported preproteins accumulated at the TOM complex under constitutive conditions, whereas the mitochondrial compromised protein import response operates during stress conditions[16,17]. Pth2 is expressed and interacts with the TOM complex under constitutive conditions (Fig. 3c and Extended Data Fig. 9b) like Ubx2, a core component of mitoTAD[17]. Ubx2 binds to TOM and recruits the AAA-ATPase Cdc48, thereby promoting the release of translocation-arrested preproteins from TOM[17]. A yeast double mutant lacking both *PTH2* and *UBX2* showed a synthetic growth defect and accumulation of Mdj1 precursor (Extended Data Fig. 10d,e), and tagged Ubx2 co-purified with TOM subunits and Pth2[17] (Fig. 5a). To test whether Pth2 cooperates with Ubx2 in TOM binding, we analysed the association of Ubx2 and Pth2 with the TOM complex in the reciprocal single-deletion mutants. The affinity purifications showed that the TOM–Ubx2 interaction does not depend on the presence of Pth2, and the TOM–Pth2 interaction does not depend on Ubx2; moreover, recruitment of Cdc48 to the mitochondria and the TOM complex occurred independently of Pth2 (Fig. 5b,c and Extended Data Fig. 10). Thus, Pth2 and Ubx2–Cdc48 bind to the TOM complex independently of each other.

Fig. 5: The relationship of Pth2 to the quality-control factors Ubx2, Vms1 and Dsk2.



a, WT and Ubx2–HA mitochondria were lysed with digitonin and affinity-purified. Proteins were analysed using SDS–PAGE and immunodetection. Load 1%, elution 100%. **b**,**c**, WT, Tom40–HA and Tom40–HA *pth2*Δ (**b**) and Tom40–HA *ubx2*Δ (**c**) cell extracts were lysed with digitonin and affinity-purified. Proteins were analysed using SDS–PAGE and immunodetection. Load, 0.2%; elution, 100%. The asterisk marks an unspecific signal of anti-Ubp16. **d**, Serial dilutions of the indicated strains were spotted onto selective medium with glycerol as a carbon source and grown at 37 °C. EV, empty vector. **e**, Cartoon of the linear structure of Pth2. TM domain, transmembrane domain (amino acids 12–32). The amino acid exchange in the Pth2-mutant form is indicated. Cellular fractionation of the indicated strains. Post-nuclear supernatant (PNS) fractions enriched for mitochondria (P13) and the cytosol (S100) were analysed by

immunodetection. Phosphoglycerate kinase 1 (Pgk1) was the cytosolic control. **f**, Serial dilutions of the indicated strains were spotted onto full medium with either glucose or glycerol as a carbon source. **g**, Cell extracts of the indicated strains were analysed by immunodetection. p, precursor; m, mature form of Mdj1. **h**, The proposed model of constitutive quality control for the removal of precursor proteins accumulated at the mitochondrial protein entry gate. Rsp5 ubiquitylates precursor proteins accumulated at the TOM complex. Ubp16 removes faulty ubiquitins from precursor proteins to enable their transport into the mitochondria. Ubx2 interacts with the TOM complex and cooperates with the AAA-ATPase Cdc48 in the transfer of ubiquitylated precursor proteins to the proteasome[17]. Pth2 binds to the TOM complex independently of Ubx2 and cooperates with Dsk2 in the transfer of precursor proteins to the proteasome. OM, outer membrane.

Second, Vms1 is a ribosome-binding cytosolic peptidyl-tRNA hydrolase that facilitates the release and diminishes the aggregation of ribosome-stalled polypeptides during co-translational mitochondrial import[36,37,38]. A double deletion of *VMS1* and *PTH2* led to a strong synthetic growth defect (Extended Data Fig. 10d). The expression of a Pth2 mutant form with an inactive peptidyl-tRNA hydrolase (Pth2(D174A))[24,37] did not rescue the growth of *pth2*Δ*vms1*Δ cells; however, a Pth2 mutant lacking the single mitochondrial outer-membrane anchor at the amino terminus (Pth2(ΔTM)) was able to rescue *pth2*Δ*vms1*Δ cells (Fig. 5d,e). The mitochondrial localization of Pth2 is therefore not needed for its function as a peptidyl-tRNA hydrolase in the Vms1 pathway. By contrast, Pth2(D174A) but not Pth2(ΔTM) rescued the growth of *pth2*Δ*pam17*Δ cells (Fig. 5d), revealing that the role of Pth2 in the removal of accumulated preproteins requires its anchoring in the outer membrane. Thus, Pth2 functions in two different experimentally separable quality-control processes. Its peptidyl-tRNA hydrolase activity is required for Vms1-linked quality control at ribosome-stalled polypeptides, whereas the mitochondrial localization of Pth2 is needed for the removal of non-imported preproteins.

Third, Pth2 interacts with the nuclear and cytosolic ubiquitin-binding protein Dsk2 (dominant suppressor of *kar1*), a ubiquitin-like-domain-containing protein that shuttles ubiquitylated substrates to the proteasome[24,39]. The human homologues of Dsk2, ubiquilins, can bind to proteins with

transmembrane segments, including non-imported mitochondrial outer membrane precursors, in the cytosol and deliver them for proteasomal degradation[40]. The identification of the TOM-linked Pth2 pathway raised the possibility that Dsk2 may participate in mitochondria-located quality control. Mutant cells lacking Dsk2 indeed displayed growth defects under respiratory conditions, particularly at elevated temperatures when high mitochondrial activity is required (Fig. 5f). The precursor form of Mdj1 accumulated in *dsk2*Δ*pam17*Δ cells (Fig. 5g). Furthermore, increased amounts of ubiquitylated preproteins were detected in the absence of Pth2, the mitochondrial receptor of Dsk2[24] (Figs. 3e and 4g), indicating that Dsk2 is involved in the removal of non-imported ubiquitylated mitochondrial preproteins.

Taken together, the MitCOM analysis of TOM interactors uncovered factors for ubiquitylation/deubiquitylation at the mitochondrial import site and a Pth2–Dsk2 pathway for the removal of accumulated preproteins (Fig. 5h).

Conclusions

MitCOM is a comprehensive high-resolution complexome of mitochondria. It covers more than 900 high-confidence yeast mitochondrial proteins and identifies more than 5,200 distinct protein peaks, revealing a considerable complexity of mitochondrial protein assemblies ranging from stable complexes to supercomplexes, assembly intermediates and dynamic interactors. The exquisite resolution and quantitative nature of the complexome enable a precise comparison of the shapes of co-migrating proteins and therefore efficient identification of protein interactors that are subsequently verified by biochemical and functional assays. The vast majority of mitochondrial proteins display multiple peaks in their abundance–mass profiles, with an average of six peaks per protein, indicating that interactions with more than one partner is a widespread characteristic. Owing to the stringency of our evaluation procedure, the peak numbers reported here probably underestimate the complexity of mitochondrial assemblies.

We used mitochondrial membrane protein complexes and quality control to illustrate the power of MitCOM for revealing interactions and pathways. (1)

We identified the direct link of the Om14 receptor for cytosolic ribosomes[15] to the mitochondrial protein import site by identifying two different pools of Om14 in the mitochondrial outer membrane, one of them directly interacting with the TOM complex. (2) We identified the yeast J-protein partner Mdj2 of the prohibitin scaffold–AAA protease megacomplex of the mitochondrial inner membrane. (3) The identification of the peptidyl-tRNA hydrolase Pth2 at the TOM complex led us to identifying a quality-control pathway that interacts with TOM independently of the established mitoTAD pathway[17]. Together with a targeted screening for E3 ubiquitin ligases and deubiquitylases, this led to the identification of quality-control mechanisms that operate at the TOM complex (Fig. 5h). First, the multifunctional cytosolic E3 ligase Rsp5[29,31] ubiquitylates mitochondrial preproteins to facilitate their degradation. Second, Ubp16—a predicted outer-membrane deubiquitylase of previously unknown function—modulates preprotein quality control by removing ubiquitin moieties. Third, Pth2 constitutes an alternative degradation pathway by cooperating with ubiquilins[24,40] in the removal of accumulated preproteins. Thus, under constitutive conditions, protein entry into the mitochondria is controlled by coordinated ubiquitylation and deubiquitylation and at least two distinct transfer modes of non-imported preproteins to the proteasome for degradation. The quality-control factors are considerably less abundant than TOM[6] (Extended Data Fig. 9a), indicating that they transiently interact with the translocase. This underscores the high sensitivity of the MitCOM approach in identifying dynamic protein interactions of low abundance.

A majority of processes and mechanisms of mitochondrial biogenesis, function and quality control have been conserved in evolution from yeast to humans, including numerous proteins that are linked to mitochondrial diseases[1,3,20]. The large number of assemblies and interactions resolved by MitCOM suggests that many mitochondrial functions and pathways await their assignment, for physiological as well as pathophysiological processes.

Methods

Yeast strains and growth conditions

The yeast strains used in this study are derived from *S. cerevisiae* BY4741 and YPH499 (Supplementary Table 4). The wild-type strain YPH499 as well as the mutant strains *Cox4*$_{His}$, *Tom20*$_{His}$, *Tom40*$_{HA}$, *Tom40*$_{HA}$ *ubx2*Δ, *ubx2*Δ, *pam17*Δ and *pre9*Δ*pam17*Δ were described previously[17,42,43]. The wild-type strain BY4741 as well as the mutant strains *pth2*Δ, *ubx2*Δ, *ubp16*Δ, *rsp5-1*, *mdm30*Δ, *mfb1*Δ and *vms1*Δ were obtained from EUROSCARF. Tagging and deletion of open reading frames were performed by homologous recombination using DNA cassettes amplified by PCR using Taq and Vent polymerase (NEB)[44] or KOD hot-start DNA-Polymerase (Merck Millipore). The genetic information for a triple HA tag was chromosomally introduced before the STOP codon of the open reading frame of *TOM40* in *pth2*Δ and *tom70*Δ cells using the cassette amplified from pFA6a 3HA-His3MX6[45]. The open reading frame of *DSK2* was deleted using the pFA6a KanMX4 cassette[46]. Deletion of *PTH2* in *ubp16*Δ and *ubx2*Δ, *VMS1* in *pth2*Δ and *UBP16* in *rsp5-1* was performed using the pFA6A His3MX6 cassette[46]. Deletion of *PAM17* in the *pth2*Δ, *dsk2*Δ, *rsp5-1*, *ubp16*Δ, *mfb1*Δ, *mdm30*Δ, *hrd1*Δ and *doa10*Δ strains, and deletion of *TOM70* in the BY4741 and Tom40$_{HA}$ strains was performed using the pFA6a hphNT1 cassette[44] (the deletion cassettes used are shown in Supplementary Table 5). Yeast strains were cultured according to standard protocols at temperatures between 24 °C and 37 °C in complete YP-medium (1% (w/v) yeast extract, 2% (w/v) bacto-peptone) or selective minimal medium (SM) (0.67% (w/v) yeast nitrogen base with ammonium sulfate; 0.07% (w/v) amino acid mixture) containing 2% (w/v) glucose (YPD, SMD), 2% (w/v) sucrose (YPS, SMS), 2% (w/v) galactose (YPGal, SMGal) or 3% (w/v) glycerol as carbon source. The cell cultures were grown until the early logarithmic growth phase, on the basis of the optical density at a wavelength of 600 nm (OD_{600}).

Construction of plasmids

Pth2 was cloned with its endogenous promoter (962 bp upstream of the start codon) into pRS416[43]. The Pth2 D174A and b_2-DHFR^{GGxY} L251G/P252G mutations were generated by site-directed mutagenesis. The putative transmembrane domain of Pth2 (amino acids 12–32) was deleted by PCR amplification of the entire plasmid without the region encoding amino acids 12–32 followed by in vitro recombination using HiFi assembly (NEB). A list

of the plasmids used is provided in Supplementary Table 5. Plasmids were used for the expression of cytochrome b_2 precursor variants containing DHFR. Folding of the DHFR domain prevents the complete translocation of the preproteins through the TOM channel and therefore arrests the N-terminal b_2-part in the mitochondrial import site. Two types of b_2-DHFR precursors were accumulated in the import site—$b_2\Delta$-DHFR, which carries a matrix-targeting signal, and b_2-DHFR, which contains both a matrix-targeting signal and an inner membrane sorting signal[17,47].

Growth analysis of yeast strains

For comparing the growth of different yeast strains, exponentially growing cells were diluted to an OD_{600} of 1.0 and diluted 1:5 (five times). The dilutions were spotted onto YP or SM plates containing glucose or glycerol as the sole carbon source and incubated at the indicated temperatures. Pictures of plates were taken after 1 to 4 days, depending on growth temperature and carbon source.

Isolation of mitochondria

Purification of mitochondria was performed by differential centrifugation[48]. Yeast cells were collected at an early logarithmic growth phase (5,500*g*, 8 min, 24 °C). Cells were washed with distilled H_2O and resuspended in DTT buffer (100 mM Tris/HCl pH 9.4, 10 mM dithiothreitol (DTT)) at a concentration of 2 ml per g wet weight of the cell pellet, followed by incubation for 30–45 min at growth temperature under constant shaking. Cells were then washed in zymolyase buffer (1.2 M sorbitol, 20 mM KP_i pH 7,4) and resuspended in zymolyase buffer at a concentration of 7 ml per g of cells. Subsequently, cells were incubated with 4 mg zymolyase per g of cells under constant shaking for 30–45 min at growth temperature to digest the cell wall. Next, cells were pelleted (2,500*g*, 5 min, 24 °C) and washed once with zymolyase buffer. The obtained spheroplasts were resuspended in ice-cold homogenization buffer (0.6 M sorbitol, 10 mM Tris/HCl pH 7.4, 1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM phenylmethylsulfonylfluoride (PMSF), 0.2% (w/v) bovine serum albumin) using 6.5 ml of buffer per g of cells. Cells were homogenized using a glass

potter with 15 strokes up and down. Subsequently, cell debris and large organelles like the nucleus were removed (2,500*g*, 5 min, 4 °C). The supernatant was centrifuged to isolate the mitochondria (17,000*g*, 15 min, 4 °C). The mitochondrial pellet was resuspended in SEM buffer (250 mM sucrose, 10 mM MOPS/KOH pH 7.2, 1 mM EDTA) and washed again in SEM buffer. The isolated mitochondria were resuspended in SEM buffer. The protein concentration was determined using the Bradford assay and mitochondria were aliquoted at a protein concentration of 10 mg ml^{-1}. Mitochondria were frozen in liquid nitrogen and stored at −80 °C.

Cryo-slicing blue native gel electrophoresis

For high-resolution complexome profiling, a blue native gradient gel (2–13% (w/v) acrylamide, 0.06–0.40% (w/v) bis-acrylamide, 67 mM ε-amino *n*-caproic acid, 50 mM Bis-Tris/HCl, pH 7.0) was used. Mitochondria corresponding to 1 mg protein amount were pelleted and solubilized in 0.8 ml lysis buffer (20 mM Tris/HCl pH 7.4, 0.1 mM EDTA, 50 mM NaCl, 10% (v/v) glycerol) containing 1% (w/v) digitonin for 30 min on ice (we used optimized conditions for a mild and efficient lysis of the yeast mitochondrial preparation by applying 8 mg purified digitonin to 1 mg mitochondrial protein; the protein to digitonin ratio of 1:8 enables efficient extraction of yeast mitochondrial membrane protein complexes, yet is milder than the 1:10 protein:digitonin ratio that is often used for yeast mitochondria[49] or the application of other detergents[8]; this is illustrated by the predominant presence of the physiological, fully assembled dimer of the F_1F_O-ATP synthase in comparison to the monomer (Fig. 1b and Extended Data Fig. 8d), whereas the typical 1:10 ratio conditions lead to an about equal distribution between dimer and monomer on blue native gels[49]). Subsequently, the sample was loaded onto a sucrose gradient consisting of 50% (w/v) sucrose and 20% (w/v) sucrose. After centrifugation, the upper phase was removed and the remaining supernatant was mixed with loading dye (0.5% (w/v) Coomassie G-250, 50 mM ε-amino *n*-caproic acid, 10 mM Bis-Tris/HCl, pH 7.0). The sample was applied to a loading zone of 5 cm width. Electrophoresis was performed at 15 mA in the presence of BN cathode buffer (0.02% (w/v) Coomassie G-250, 50 mM Tricine, 15 mM Bis-Tris/HCl, pH 7.0) and anode buffer (50 mM Bis-Tris/HCl, pH 7.0). After 1 h, the BN cathode buffer was replaced by a cathode buffer lacking Coomassie

G-250 and the electrophoresis was continued for 2.5 h at 15 mA. A 2.5 cm lane was then excised, fixed in 30% ethanol/15% acetic acid, embedded in tissue embedding medium (Leica) and subjected to cryo-slicing[50]. Using a step size of 0.3 mm along the gel lane, 245 slices were obtained, extensively washed and separately digested with trypsin[50].

MS analysis

The trypsin-digested peptides were dissolved in 20 µl sample buffer (0.5% (v/v) trifluoroacetic acid in H_2O) and 1 µl aliquots (or less) were taken for LC–MS/MS analysis. Loading onto the precolumn (PepMap 100, C18 stationary phase) was achieved through an autosampler of a split-free UltiMate 3000 RSLCnano HPLC (Dionex/Thermo Fisher Scientific). Subsequent elution and separation on the SilicaTip column emitter (inner diameter, 75 µm; tip, 8 µm; New Objective,; packed 23 cm with ReproSil-Pur 120 ODS-3 (C18 stationary phase; Dr. Maisch HPLC)) occurred during a three-step linear gradient generated from eluent A (0.5% (v/v) acetic acid) and eluent B (0.5% (v/v) acetic acid in 80% (v/v) acetonitrile): after 5 min equilibration in 3% B, 90 min from 3% B to 30% B; 20 min from 30% B to 50% B; and 10 min from 50% B to 99% B. Subsequent column washing/regeneration comprised 5 min 99% B; 5 min from 99% B to 3% B; 10 min 3% B. The flow rate was set to 300 nl min^{-1}. Electrospray parameters were positive ion mode, spray voltage 2.3 kV, transfer capillary temperature 300 °C. Data were acquired on the QExactive HF-X mass spectrometer (Thermo Fisher Scientific) with the following settings: maximum MS/MS injection time = 200 ms; dynamic exclusion time = 45 s; minimum signal intensity threshold = 40,000 (counts), fragmentation = 15 top precursors; mass isolation width = 1.0 *m*/*z*.

Protein identification

Primary MS data were processed using msconvert (https://proteowizard.sourceforge.io; v.3.0.11098; settings: Mascot generic format, filter options ‘peakPicking true 1’, ‘threshold count 500 most-intense’). Obtained peak lists were then *m*/*z*-calibrated using the MaxQuant raw file processor (v.1.6.17, https://www.maxquant.org)[51,52] and used as input for a Mascot (v.2.7, Matrix Science) database search against the

UniProtKB/SwissProt yeast database (SwissProt_YEAST_20201007) with general contaminations added (GPM cRAP database; cRAP_20190304). Match parameters were as follows: precursor mass tolerance = ±2.5 ppm, variable modifications = acetyl (protein N-term), carbamidomethyl (C), formyl (N-term, S, T), Gln->pyro-Glu (N-term Q), Glu->pyro-Glu (N-term E) and oxidation (M), fragment mass tolerance = ± 20 mmu, missed tryptic cleavage(s) = 1. Export filter settings were as follows: peptide-spectrum-match (PSM) FDR = 3%, minimum ion score = 0.5, grouping of related protein hits used the name of the predominant member. Exogenous contaminants (for example, keratins, trypsin, IgG chains) or protein identifications based on only one specific peptide in less than three slice samples were not considered further.

Protein quantification

For label-free quantification of proteins, LC–MS data were processed as previously described[9,50,53,54] with some crucial improvements. MaxQuant (v.1.6.17; https://www.maxquant.org)[51,52] was used to determine and mass-calibrate peptide signal intensities (peak volumes) from recorded FT full scans. Systematic variations in peptide elution times were corrected by LOESS regression after pairwise alignment of the datasets (median peptide elution times over all of the aligned datasets were used as a reference). Matching of peak volumes and peptide identities (obtained either directly or indirectly from MS/MS database matches) was achieved using custom developed software[53] (*m*/*z* and elution time matching tolerances were ±1.5–2 ppm and ±1 min, respectively). Global offsets in peptide intensity between runs were corrected by normalization to the local median of the relative peptide intensities (consistently assigned peptides within a window of 40 slices). For effectively eliminating the impact of missing, non-consistent or incorrectly assigned peptide intensity values, we applied an additional procedure consisting of four steps. First, the accuracy of all assigned peptide intensity values was determined by analysing matrices (peptides versus runs) of protein-specific peptide intensity values for their internal consistency. Within these matrices, the pairwise relationships of peptide intensity values between and within MS-runs (in all possible combinations) provided distributions of predicted intensities for connected matrix cells from which expected intensity values (EPVs) could be calculated (by kernel density

estimation), which served as measure of consistency of the respective peak volume values and were later used as weighting factors. If insufficient data precluded determination of EPVs, values interpolated from EPV-validated intensities in neighbouring (window of 5) slices/datasets were used alternatively. Second, a time- and run-dependent detectability threshold was estimated for each of the matrix cells (peptides versus runs) using the third percentile of intensity values from peptides co-eluting within a 3 min time window. Third, for each protein, peptide intensity values from qualified runs (that is, consistent protein-specific peak volume values of respective columns in each protein matrix) were merged (EPV-weighted least squares fits to the dataset with the highest number of peptide intensities assigned to the respective protein) into a single intensity value vector, termed protein reference ridge. These vectors reflected the maximum protein coverage of MS/MS-identified and quantified peptides with their relative ionization efficiencies and were used to determine molecular abundances ($\text{abundance}_{\text{norm}}$ spec values) as described previously[53]. Fourth, protein quantification was achieved by a weighted fitting of its measured peptide intensities in five consecutive slices (equivalent to a sliding average) to its reference ridge (Extended Data Fig. 2). Quantification details for each data point in a protein profile (Supplementary Table 2) are provided in the 'peptide details' feature of the expert viewer tool (Extended Data Fig. 3; https://www.complexomics.org/datasets/mitcom).

The oversampling of the blue native gel separation (0.3 mm step size) was used to provide robust protein quantification without compromising effective size resolution. Thus, each protein abundance value, that is, each data point in an abundance mass profile, integrated the results from five consecutive LC–MS/MS analyses. Moreover, the processing of LC–MS input data described above provided objective measures for reliability and accuracy of protein quantification: (1) the number of protein-specific peptide intensities used and (2) the deviation of these peptide intensities from the expected peptide intensities. This information was integrated into a 'reliability score' (ranging from 0 (not significant) to 1 (maximum reliability)) for each protein abundance value (Supplementary Table 2).

Evaluation, accession and visualization of data

A total of 1,891 proteins with a required minimal number of two protein-specific peptides were identified in the entire csBN–MS dataset. Of these, 906 proteins were considered to be bona fide mitochondrial proteins on the basis of respective annotations in either the UniProtKB/SwissProt database, the Yeast Genome Database (SGB) or ref. [6]. Together, the 906 mitochondrial proteins represent the core of the complexome dataset termed the MitCOM (Supplementary Tables 1 and 2). Information on (1) molecular mass and membrane-spanning helices was inferred from the UniProtKB/SwissProt database, (2) localization in submitochondrial compartments was extracted from the respective SGD GO terms, and (3) functional classification was taken from ref. [6] supplemented by GO annotations and literature for the proteins not already classified therein (Fig. 2, Supplementary Table 1 and Extended Data Fig. 6). Among the proteins identified with one protein-specific peptide only, an additional 49 mitochondrial proteins were found to be potentially significant and were separately added to the list of MitCOM proteins in Supplementary Table 1. The remaining 985 proteins, all identified by at least two protein-specific peptides, were classified as non-mitochondrial proteins predominantly localized in the endoplasmic reticulum and cytosol (Fig. 1, Extended Data Fig. 7 and Supplementary Table 3). Information on their molecular mass and membrane integration, their preferred subcellular localization and their functional classification was obtained as described for the mitochondrial proteins above.

The principles of mass estimation of protein complexes using blue native electrophoresis were outlined in ref. [55]. For converting slice numbers to apparent molecular masses of proteins/complexes we used a set of marker proteins/complexes with a defined migration pattern (that is, sharply focused profile peaks) and molecular masses broadly distributed over the sampled blue native gel range (name/subunits with references; predicted molecular mass, log[molecular mass], slice maximum): oxoglutarate dehydrogenase/ketoglutarate dehydrogenase complex (Kgd1/Odo1–Kgd2/Odo2–Lpd1/Dldh–Kgd4[56,57,58,59,60]; 3,020 kDa, 3.48, 21.4); pre-60S ribosome large subunit (RL3, RLP24, NOG1; PDB: 3JCT (ref. [61]); 2,680 kDa, 3.43, 11.5); dimer of F_1F_O-ATP synthase (complex V dimer[8]; 1,250 kDa, 3.10, 54.1); respiratory III_2IV_2 supercomplex (ref. [8]; 1,000 kDa, 3.00, 62.4); respiratory III_2IV_1 supercomplex (ref. [8]; 750 kDa, 2.88, 71.6);

monomer of F_1F_O-ATP synthase (complex V monomer[8]; 600 kDa, 2.78, 85,6); SAM–Mdm10 complex (PDB: 7BTX (ref. [62]); 185.5 kDa, 2.27, 143.8); TIM22 complex (PDB: 6LO8 (ref. [63]); 174 kDa, 2.24, 136.2); ATM1 (ABC transporter mitochondrial; PDB: 4MYC (ref. [64]); 155 kDa, 2.19, 157); SAM_{core} complex (Sam50, Sam37, Sam35[62,65]; 129.4 kDa, 2.11, 173.5); succinate dehydrogenase (complex II[8]; 130 kDa, 2.11, 182); NADP-cytochrome P450 reductase (Ncp1/NCPR[66]; 77 kDa, 1.88, 239). The resulting relationship of molecular mass over slice numbers was fitted with a sigmoidal function (IGOR Pro 9 WaveMetrics) and the resulting fit-line was used for calibration (Extended Data Fig. 2d).

Abundance–mass profiles of all MitCOM proteins (Supplementary Table 2) were analysed for their composition of individual components (peaks) using custom-developed software (complexomics-mitcom v.1.0; released as a Python package under the MIT license, available at Zenodo (https://doi.org/10.5281/zenodo.7355040)). First, locations of apparent peaks were determined by local maxima search with subsequent filtering (minimum relative height 0.1, minimum relative prominence 0.5, maximum width 50). Then, a multicomponent Gaussian model was initialized with the number and locations of the identified peaks. The model was iteratively adjusted and fitted to the profile. Up to 12 Gaussian components were added preferably at locations with large residuals, resulting in an improved fitting of highly overlapping peaks, manifesting as ‘shoulders’. Sensible limits and stop conditions were applied to avoid overfitting. A total of 818 MitCOM proteins was accessible to automated analysis. From a total number of 5,224 peaks (manually curated), 4,070 peaks were adequately fitted by our algorithm providing parameters for their apparent mass, half-width and molecular abundance (Fig. 1c). The total peak count (manually curated) was used for statistical analysis in Fig. 2 and Extended Data Fig. 6. For the non-mitochondrial proteins, the number of apparent peaks was counted manually (Extended Data Fig. 7).

Abundance–mass profiles of all MitCOM proteins were deposited in the openly accessible interactive resource platform Complexome Profiling Data Resource (CEDAR)[67], where they can be accessed through an interactive online visualization tool (https://www3.cmbi.umcn.nl/cedar/browse/experiments/crx36). Protein

profile normalization, filtering, baseline subtraction and magnification/scaling for convenient display and evaluation are integrated functions of this viewer (Extended Data Fig. 3). Moreover, using custom Pearson correlation analysis, it offers a basic method to search for proteins with similar abundance profile peaks or patterns. An extended version of the viewer with additional features enabling the inspection of the peptide intensity information underlying each datapoint of the protein profiles (Extended Data Fig. 3), an integrated help function explaining the use of the available features and a supplemental viewer containing abundance–mass profiles of the quantified non-mitochondrial proteins (that passed a quality check) are available online (https://www.complexomics.org/datasets/mitcom).

For a global view on the complexome organization of MitCOM (Extended Data Figs. 4 and 5), unsupervised protein profile matching was performed as follows (similar to the approach in ref. [11]). Successfully fitted protein profile components (Fig. 1c) were filtered by width (max 20 s.d.) and location (peak full width at half maximum fully inside gel slice range) and clipped at 2.5 s.d. These defined the boundaries of 3,263 profile segments as seed regions of interest (ROIs). Comparison of each ROI to its corresponding segment in other profiles using Pearson correlation yielded 82,268 high-correlating segments ($r \geq 0.95$) as additional ROIs. Finally, out of the total of 85,531 ROIs, the ones that originated from the same profile and had similar boundaries (within 3 slices) were merged. The resulting 49,112 ROIs were used to assess similarity of protein components. To this end, a distance metric was designed that incorporated the following ROI-specific values: (1) slice index of left boundary; (2) slice index of right boundary; (3) maximum abundance; (4) r value from correlation with seed ROI (average if ROIs were merged). All of the values were minimum/maximum-normalized, except for boundaries, which were square-root-transformed, giving the highest weight to component locations and quickly penalizing differences. The distance of any two ROIs is determined by the Euclidean distance of their respective value vectors and the coefficient r of their mutual correlation. With the maximum theoretical distance being a dataset-specific fixed value, distances could easily be converted to a normalized similarity score ranging from 0 (most dissimilar) to 1 (identical location and abundance values). On the basis of the custom distance metric, pairwise

distances of all protein ROIs were calculated and used for visualizing component similarity using a *t*-SNE plot (Extended Data Fig. 5).

Preparation of yeast cell extracts

Whole yeast cell lysates were obtained by post-alkaline extraction[68]. Exponentially growing yeast cells (OD_{600} of 2.5) were pelleted (3,000*g*, 5 min, 20 °C) and resuspended in distilled H_2O. Subsequently, the samples were mixed with the same volume of 0.2 M NaOH and incubated for 5 min at 24 °C. Cells were pelleted (3,000*g*, 5 min, 4 °C), resuspended in sample buffer (8 M urea, 5% (w/v) SDS, 1 mM EDTA, 1.5% (w/v) DTT, 0.025% (w/v) bromophenol blue, 200 mM Tris/HCl pH 6.8) and denatured for 10 min at 65 °C shaking at 1,400 rpm.

Yeast extracts for large-scale affinity purification were generated by collecting cells in the early logarithmic growth phase (5,500*g*, 8 min, 24 °C). Cells were then washed with distilled H_2O lysis buffer (0.1 M EDTA, 50 mM NaCl, 10% (v/v) glycerol, 20 mM Tris/HCl pH 7.4). Cells resuspended in lysis buffer were then frozen in liquid nitrogen and cell disruption was achieved by cryo-grinding at 25 Hz for 10 min in a Cryo Mill (Retsch). The obtained lysates were stored at −80 °C until further use.

To obtain small amounts of cell extracts for affinity purifications, cells from exponentially growing cells (OD_{600} of 100–200) were resuspended in lysis buffer with protease inhibitors (1 mM PMSF, 1× HALT protease inhibitor cocktail (Thermo Fisher Scientific)). Cells were then ruptured with silica beads 6 times 30 s with a 1 min break in between at 4 °C on a cell disruptor (Vortex Disruptor Genie). Cell extracts were cleared by centrifugation (2,000*g*, 5 min, 4 °C) and directly further processed for affinity purification.

Affinity purification of tagged proteins

For affinity purification of His-tagged proteins[17,42,69], isolated mitochondria were solubilized in lysis buffer (20 mM Tris pH 7.4, 0.1 mM EDTA, 50 mM NaCl, 10% (v/v) glycerol, 2 mM PMSF, 1× protease inhibitor cocktail without EDTA) containing 1% (w/v) digitonin and 10 mM imidazole and incubated for 15 min at 4 °C. The solubilized sample was cleared by

centrifugation (10 min, 17,000*g*, 4 °C) and incubated with Ni-NTA agarose beads (Qiagen) that were pre-equilibrated in lysis buffer with 0.1% (w/v) digitonin and 10 mM imidazole. After 1 h incubation at 4 °C under constant rotation, unbound proteins were removed and the affinity matrix was washed with excess amount of lysis buffer containing 0.1% (w/v) digitonin and 40 mM imidazole. Bound proteins were eluted with lysis buffer containing 0.1% (w/v) digitonin and 250 mM imidazole. After addition of sample buffer proteins were denatured for 10 min at 60 °C (Cox4–His) or 5 min at 96 °C.

For affinity purification of HA-tagged proteins[17,69], isolated mitochondria or cell extracts were solubilized in lysis buffer containing 1% (w/v) digitonin. After solubilization for 15 min (purified mitochondria) or 30 min (cell extracts), the samples were cleared by centrifugation (17,000*g*, 10 min, 4 °C). The supernatant was incubated for 1 h at 4 °C under constant rotation with an anti-HA affinity matrix (Roche) that was pre-equilibrated with 0.5 M acetate followed by washing with lysis buffer containing 0.1% (w/v) digitonin. Unbound proteins were removed and the beads were washed with an excess amount of lysis buffer containing 0.1% (w/v) digitonin. Proteins were eluted by incubation with sample buffer at 95 °C.

For purification of Phb1–protein A, mitochondria were solubilized in lysis buffer containing 1% (w/v) digitonin for 1 h at 4 °C. The lysate was cleared by centrifugation (17,000*g*, 10 min, 4 °C) and incubated for 1.5 h at 4 °C with IgG Sepharose (Cytiva). The IgG Sepharose was washed 10 times with wash buffer (20 mM Tris-HCl pH 7.4, 60 mM NaCl, 10% (v/v) glycerol, 0.1 mM EDTA, 2 mM PMSF, 0.3% (w/v) digitonin). Phb1 and bound proteins were eluted by cleavage of the protein A tag using TEV protease (Thermo Fisher Scientific) in wash buffer at 24 °C for 2.5 h. Subsequently, the protease was removed through its His tag by incubating the eluates with equilibrated Ni-NTA for 30 min at 4 °C.

For tandem purification of the TOM complex[70,71] through Tom22–His and Tom40–Strep, isolated mitochondria were solubilized in tandem buffer (100 mM Tris/HCl pH 8.0, 150 mM NaCl, 10% (v/v) glycerol, 1× protease inhibitor cocktail) supplemented with 5 mM imidazole and 3% (w/v) digitonin for 1 h rotating. The samples were cleared (17,000*g*, 10 min, 4 °C) and the subsequent purification steps were performed using the ÄKTA

Explorer 100 system (Cytiva). In the first steps, Tom22–His was purified using the HisTrap HP column pre-equilibrated with tandem buffer containing 5 mM imidazole and 0.1% (w/v) digitonin. The column was washed with tandem buffer containing 5 mM imidazole and 0.1% (w/v) digitonin and bound proteins were eluted with tandem buffer containing 250 mM imidazole and 0.1% (w/v) digitonin. In the second purification step via Tom40–Strep, the elution sample was applied onto a Strep-Tactin HP column pre-equilibrated with tandem buffer containing 0.1% (w/v) digitonin. The column was washed with excess amount of tandem buffer containing 0.1% (w/v) digitonin and bound protein was eluted with tandem buffer containing 0.1% (w/v) digitonin and 10 mM biotin.

Purification of ubiquitin-modified proteins

Proteins conjugated to His-tagged ubiquitin were purified through Ni-NTA agarose under denaturing conditions[41,72]. Cells expressing His-tagged ubiquitin were grown to logarithmic growth phase. Cells corresponding to an OD_{600} of 200 were collected (2,500*g,* 4 min, 4 °C) and washed with distilled H_2O. Cells were resuspended under denaturing conditions to inactivate proteases including deubiquitylating enzymes in 1 ml buffer A (6 M guanidinium hydrochloride, 100 mM NaH_2PO_4, 10 mM Tris-HCl, pH 8.0). Silica beads (diameter 0.5 mm) were added to the samples and cells were disrupted using a cell disruptor (Vortex Disruptor Genie). Cellular lysates were cleared (500*g*, 5 min, 4 °C) to remove residual beads. Subsequently, the samples were diluted 1:10 in the presence of 0.05% (v/v) Tween-20 and incubated for 1 h at 24 °C under constant rotation. Insoluble material was removed (3,500*g*; 10 min; 4 °C) and the remaining supernatants were incubated in the presence of 20 mM imidazole with Ni-NTA agarose beads overnight at 4 °C under constant rotation. After removal of unbound proteins, the beads were washed twice with buffer A containing 20 mM imidazole and 0.05% Tween-20, followed by five washing steps with buffer C (8 M urea, 100 mM NaH_2PO_4, 10 mM Tris-HCl, pH 6.3, 0.05% Tween-20, 20 mM imidazole). Bound proteins were eluted with HU sample buffer (8 M urea, 5% (w/v) SDS, 1 mM EDTA, 1.5% (w/v) DTT, 0.025% (w/v) bromophenol blue, 200 mM Tris-HCl pH 6.8) for 10 min at 65 °C.

Subcellular fractionation

Cells were fractionated by differential centrifugation[17]. Cells were grown to an early logarithmic growth phase. Cells corresponding to an OD_{600} of 100 were resuspended in DTT buffer (100 mM Tris pH 9.4, 10 mM DTT) and incubated at 30 °C (20 min, 900 rpm). Cells were then resuspended in zymolyase buffer (1.2 M sorbitol, 20 mM KP_i pH 7.4), zymolyase was added at a final concentration of 100 mg ml^{-1} and cells were incubated at 30 °C (45 min, 900 rpm). Cells were washed with zymolyase buffer and resuspended in ice cold homogenization buffer (0.6 M sorbitol, 10 mM Tris/HCl pH 7.4, 1 mM EDTA, 1 mM PMSF, 0.2% (w/v) bovine serum albumin). Cells were homogenized using a glass potter with 20 strokes up and down. Subsequently, cell debris and large organelles such as the nucleus were removed (1,500*g*, 5 min, 4 °C). A fraction of the supernatant was collected as post-nuclear supernatant, the remaining supernatant was subjected to centrifugation to isolate mitochondria (17,000*g*, 15 min, 4 °C). The mitochondrial pellet was resuspended in SEM buffer (250 mM sucrose; 10 mM MOPS/KOH pH 7.2; 1 mM EDTA), mitochondria were purified twice by centrifugation through a sucrose cushion (500 mM sucrose; 10 mM MOPS/KOH pH 7.2; 1 mM EDTA) and mitochondria were resuspended in SEM buffer (P13 fraction). The supernatant of the first 17,000*g* centrifugation was ultracentrifuged (100,000*g*, 1 h, 4 °C) and the supernatant was collected as the S100 fraction.

Blue native gel electrophoresis for standard analysis

For blue native gel separation[48], mitochondria were solubilized in lysis buffer (20 mM Tris/HCl pH 7.4, 0.1 mM EDTA, 50 mM NaCl, 10% (v/v) glycerol) containing 1% (w/v) digitonin. The samples were cleared by centrifugation (17,000*g*, 15 min, 4 °C) and loading dye was added to a final concentration of 0.5% (w/v) Coomassie Brilliant Blue G-250, 10 mM Bis-Tris pH 7.0 and 50 mM 6-aminocaproic acid before loading onto the blue native gel. The prime symbols at the molecular mass markers of the blue native gels in Fig. 3b and Extended Data Fig. 10b,h indicate the correlation between the migration of water-soluble markers and membrane-protein markers according to ref. [62].

Immunoblotting

Proteins separated on polyacrylamide gels were transferred by semi-dry blotting to a polyvinylidene fluoride membrane (Milipore) with blotting buffer (20% (v/v) methanol, 150 mM glycine, 0.02% (w/v) SDS, 20 mM Tris base) for 2 h at 250 mA. After blotting, membranes were blocked with 5% (w/v) skimmed milk powder in TBS-T (12.5 mM NaCl, 20 mM Tris/HCl, pH 7.4, 0.1% (v/v) Tween-20) or RotiBlock (Roth) for 1 h. The membranes were incubated with primary antibodies for 1–2 h at room temperature or overnight at 4 °C. The membranes were washed with an excess of TBS-T and incubated with secondary antibodies at room temperature for 1 h. Different types of secondary antibodies were used. For detection of immune signals using the Licor system, secondary antibodies against rabbit or mouse were coupled to fluorescent labels (IRDye 800CW, anti-mouse; IRDye 800CW, anti-rabbit; IRDye 680RD, anti-mouse). For detection with the image analyser or X-ray films, an anti-rabbit antibody coupled to horseradish peroxidase was used. Unbound antibodies were removed by washing with excess of TBS-T. The signal of horseradish peroxidase coupled secondary antibodies was detected after incubating the membrane with enhanced chemiluminescence solution[73] using either an Amersham Imager 680 (Cytiva) or the LAS3000 image reader (FujiFilm). Fluorescent secondary antibodies were detected on Odyssey CLx Infrared Imaging System (Li-Cor) and analysed using Image Studio (v.5.2.5; Li-Cor). The specificity of the immunosignals was confirmed by their absence in cells or mitochondria from deletion strains. In case of essential genes, the size shift of the band in cells expressing tagged-proteins confirmed the specificity of the immunosignal. The separating white lanes indicate where irrelevant gel lanes were digitally removed. A list of the antibodies used is provided in Supplementary Table 6. Experiments were typically run on several gels, which were analysed in parallel by western blotting, including control western blots. Representative blots were selected and processed from the original files (Supplementary Fig. 1) using ImageJ (v.2.1.0), Adobe Photoshop 2021 and Adobe Illustrator 2021.

Reproducibility and image processing

Representative images are shown for growth and biochemical assays/western blotting, including analysis of yeast growth (wild-type and mutants), total cell extracts, affinity purification from cell extracts, subcellular fractionation, protein steady state levels, blue native electrophoresis and affinity purification from isolated mitochondria. The findings were confirmed by independent experiments for the following figures (minimum number of independent experiments in parentheses): Figs. 3b (2), 3d (3), 3e (2), 4b (2), 4c (2), 4d (2), 4e (2), 4f (2), 4g (2), 5a (2), 5b (3), 5c (2), 5d (3), 5e (2), 5f (3) and 5g (2) and Extended Data Figs. 8f (3), 8h (2), 9b (2), 9c (2), 9d (2), 9e (2), 9f (5), 9g (3), 9h (2), 10a (2), 10b (2), 10c (3), 10d (2), 10e (2), 10f (2), 10g (2) and 10h (2). Images of western blots and growth assays were processed using Adobe Photoshop 2021 and figures were assembled in Adobe Illustrator 2021.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The MS data generated during this study are available at the ProteomeXchange under identifier PXD029548. UniProtKB/Swiss-Prot (https://www.uniprot.org; SwissProt_YEAST_20201007) and *Saccharomyces* Genome Database (SGD, https://www.yeastgenome.org; GO term mapping 20201027) were used as resources for structural, cell biological and functional annotation of proteins. The MitCOM datasets, including an interactive profile viewer, are available at Complexomics (https://www.complexomics.org/datasets/mitcom) or on the CEDAR platform (https://www3.cmbi.umcn.nl/cedar/browse/experiments/CRX36). All other data are available in the figures and the Supplementary Information, including uncropped versions of gels/blots in Supplementary Fig. 1 and Supplementary Tables 1–3 (Excel files).

Code availability

Complexomics-mitcom v.1.0 (a Python package for peak detection, similarity score calculation and *t*-SNE visualization) is available at Zenodo (https://doi.org/10.5281/zenodo.7355040).

References

1. Pfanner, N., Warscheid, B. & Wiedemann, N. Mitochondrial proteins: from biogenesis to functional networks. *Nat. Rev. Mol. Cell Biol.* **20**, 267–284 (2019).

2. Nunnari, J. & Suomalainen, A. Mitochondria: in sickness and in health. *Cell* **148**, 1145–1159 (2012).

3. Morgenstern, M. et al. Quantitative high-confidence human mitochondrial proteome and its dynamics in cellular context. *Cell Metab.* **33**, 2464–2483 (2021).

4. Sung, A. Y., Floyd, B. J. & Pagliarini, D. J. Systems biochemistry approaches to defining mitochondrial protein function. *Cell Metab.* **31**, 669–678 (2020).

5. Rath, S. et al. MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. *Nucleic Acids Res.* **49**, D1541–D1547 (2021).

6. Morgenstern, M. et al. Definition of a high-confidence mitochondrial proteome at quantitative scale. *Cell Rep.* **19**, 2836–2852 (2017).

7. Wittig, I. & Malacarne, P. F. Complexome profiling: assembly and remodeling of protein complexes. *Int. J. Mol. Sci.* **22**, 7809 (2021).

8. Schägger, H. & Pfeiffer, K. Supercomplexes in the respiratory chains of yeast and mammalian mitochondria. *EMBO J.* **19**, 1777–1783 (2000).

9. Müller, C. S. et al. Cryo-slicing blue native-mass spectrometry (csBN-MS), a novel technology for high resolution complexome profiling. *Mol. Cell. Proteomics* **15**, 669–681 (2016).

10. Guerrero-Castillo, S., van Strien, J., Brandt, U. & Arnold, S. Ablation of mitochondrial DNA results in widespread remodeling of the mitochondrial complexome. *EMBO J.* **40**, e108648 (2021).

11. Nolte, H. & Langer, T. ComplexFinder: a software package for the analysis of native protein complex fractionation experiments. *Biochim. Biophys. Acta Bioenerg.* **1862**, 148444 (2021).

12. Giese, H. et al. NOVA: a software to analyze complexome profiling data. *Bioinformatics* **31**, 440–441 (2015).

13. Lauffer, S. et al. *Saccharomyces cerevisiae* porin pore forms complexes with mitochondrial outer membrane proteins Om14p and Om45p. *J. Biol. Chem.* **287**, 17447–17458 (2012).

14. Wenz, L. S. et al. The presequence pathway is involved in protein sorting to the mitochondrial outer membrane. *EMBO Rep.* **15**, 678–685 (2014).

15. Lesnik, C., Cohen, Y., Atir-Lande, A., Schuldiner, M. & Arava, Y. OM14 is a mitochondrial receptor for cytosolic ribosomes that supports co-translational import into mitochondria. *Nat. Commun.* **5**, 5711 (2014).

16. Weidberg, H. & Amon, A. MitoCPR—a surveillance pathway that protects mitochondria in response to protein import stress. *Science* **360**, eaan4146 (2018).

17. Mårtensson, C. U. et al. Mitochondrial protein translocation-associated degradation. *Nature* **569**, 679–683 (2019).

18. Wrobel, L. et al. Mistargeted mitochondrial proteins activate a proteostatic response in the cytosol. *Nature* **524**, 485–488 (2015).

19. Wang, X. & Chen, X. J. A cytosolic network suppressing mitochondria-mediated proteostatic stress and cell death. *Nature* **524**, 481–484 (2015).

20. Anderson, N. S. & Haynes, C. M. Folding the mitochondrial UPR into the integrated stress response. *Trends Cell Biol.* **30**, 428–439 (2020).

21. Doan, K. N. et al. The mitochondrial import complex MIM functions as main translocase for α-helical outer membrane proteins. *Cell Rep.* **31**, 107567 (2020).

22. Rosas-Sandoval, G. et al. Orthologs of a novel archaeal and of the bacterial peptidyl-tRNA hydrolase are nonessential in yeast. *Proc. Natl Acad. Sci. USA* **99**, 16707–16712 (2002).

23. Fromant, M., Ferri-Fioni, M. L., Plateau, P. & Blanquet, S. Peptidyl-tRNA hydrolase from *Sulfolobus solfataricus*. *Nucleic Acids Res.* **31**, 3227–3235 (2003).

24. Ishii, T., Funakoshi, M. & Kobayashi, H. Yeast Pth2 is a UBL domain-binding protein that participates in the ubiquitin-proteasome pathway. *EMBO J.* **25**, 5492–5503 (2006).

25. van der Laan, M. et al. Pam17 is required for architecture and translocation activity of the mitochondrial protein import motor. *Mol. Cell. Biol.* **25**, 7449–7458 (2005).

26. Schendzielorz, A. B. et al. Motor recruitment to the TIM23 channel's lateral gate restricts polypeptide release into the inner membrane. *Nat. Commun.* **9**, 4028 (2018).

27. Phu, L. et al. Dynamic regulation of mitochondrial import by the ubiquitin system. *Mol. Cell* **77**, 1107–1123 (2020).

28. König, T. et al. MIROs and DRP1 drive mitochondrial-derived vesicle biogenesis and promote quality control. *Nat. Cell Biol.* **23**, 1271–1286 (2021).

29. Hoppe, T. et al. Activation of a membrane-bound transcription factor by regulated ubiquitin/proteasome-dependent processing. *Cell* **102**, 577–586 (2000).

30. Kowalski, L. et al. Determinants of the cytosolic turnover of mitochondrial intermembrane space proteins. *BMC Biol.* **16**, 66 (2018).

31. Fang, N. N. et al. Rsp5/Nedd4 is the main ubiquitin ligase that targets cytosolic misfolded proteins following heat stress. *Nat. Cell Biol.* **16**, 1227–1237 (2014).

32. Goodrum, J. M., Lever, A. R., Coody, T. K., Gottschling, D. E. & Hughes, A. L. Rsp5 and Mdm30 reshape the mitochondrial network in

response to age-induced vacuole stress. *Mol. Biol. Cell* **30**, 2141–2154 (2019).

33. Kinner, A. & Kolling, R. The yeast deubiquitinating enzyme Ubp16 is anchored to the outer mitochondrial membrane. *FEBS Lett.* **549**, 135–140 (2003).

34. Kondo-Okamoto, N. et al. The novel F-box protein Mfb1p regulates mitochondrial connectivity and exhibits asymmetric localization in yeast. *Mol. Biol. Cell* **17**, 3756–3767 (2006).

35. Escobar-Henriques, M., Westermann, B. & Langer, T. Regulation of mitochondrial fusion by the F-box protein Mdm30 involves proteasome-independent turnover of Fzo1. *J. Cell Biol.* **173**, 645–650 (2006).

36. Izawa, T., Park, S. H., Zhao, L., Hartl, F. U. & Neupert, W. Cytosolic protein Vms1 links ribosome quality control to mitochondrial and cellular homeostasis. *Cell* **171**, 890–903 (2017).

37. Verma, R. et al. Vms1 and ANKZF1 peptidyl-tRNA hydrolases release nascent chains from stalled ribosomes. *Nature* **557**, 446–451 (2018).

38. Zurita Rendon, O. et al. Vms1p is a release factor for the ribosome-associated quality control complex. *Nat. Commun.* **9**, 2197 (2018).

39. Funakoshi, M., Sasaki, T., Nishimoto, T. & Kobayashi, H. Budding yeast Dsk2p is a polyubiquitin-binding protein that can interact with the proteasome. *Proc. Natl Acad. Sci. USA* **99**, 745–750 (2002).

40. Itakura, E. et al. Ubiquilins chaperone and triage mitochondrial membrane proteins for degradation. *Mol. Cell* **63**, 21–33 (2016).

41. den Brave, F. et al. Chaperone-mediated protein disaggregation triggers proteolytic clearance of intra-nuclear protein inclusions. *Cell Rep.* **31**, 107680 (2020).

42. Böttinger, L. et al. A complex of Cox4 and mitochondrial Hsp70 plays an important role in the assembly of the cytochrome *c* oxidase. *Mol. Biol. Cell* **24**, 2609–2619 (2013).

43. Sikorski, R. S. & Hieter, P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* **122**, 19–27 (1989).

44. Janke, C. et al. A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast* **21**, 947–962 (2004).

45. Knop, M. et al. Epitope tagging of yeast genes using a PCR-based strategy: more tags and improved practical routines. *Yeast* **15**, 963–972 (1999).

46. Longtine, M. S. et al. Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **14**, 953–961 (1998).

47. Boos, F. et al. Mitochondrial protein-induced stress triggers a global adaptive transcriptional programme. *Nat. Cell Biol.* **21**, 442–451 (2019).

48. Priesnitz, C., Pfanner, N. & Becker, T. Studying protein import into mitochondria. *Methods Cell Biol.* **155**, 45–79 (2020).

49. Wagner, K. et al. Mitochondrial F_1F_o-ATP synthase: the small subunits e and g associate with monomeric complexes to trigger dimerization. *J. Mol. Biol.* **392**, 855–861 (2009).

50. Müller, C. S. et al. High-resolution complexome profiling by cryoslicing BN-MS analysis. *J. Vis. Exp.* **152**, e60096 (2019).

51. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).

52. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319 (2016).

53. Bildl, W. et al. Extending the dynamic range of label-free mass spectrometric quantification of affinity purifications. *Mol. Cell. Proteomics* **11**, M111.007955 (2012).

54. Schwenk, J. et al. An ER assembly line of AMPA-receptors controls excitatory neurotransmission and its plasticity. *Neuron* **104**, 680–692 (2019).

55. Wittig, I., Beckhaus, T., Wumaier, Z., Karas, M. & Schägger, H. Mass estimation of native proteins by blue native electrophoresis and practical hints. *Mol. Cell. Proteomics* **9**, 2149–2161 (2010).

56. Repetto, B. & Tzagoloff, A. In vivo assembly of yeast mitochondrial α-ketoglutarate dehydrogenase complex. *Mol. Cell. Biol.* **11**, 3931–3939 (1991).

57. Vögtle, F. N. et al. Global analysis of the mitochondrial N-proteome identifies a processing peptidase critical for protein stability. *Cell* **139**, 428–439 (2009).

58. Heublein, M. et al. The novel component Kgd4 recruits the E3 subunit to the mitochondrial α-ketoglutarate dehydrogenase. *Mol. Biol. Cell* **25**, 3342–3349 (2014).

59. Ambrus, A. et al. Formation of reactive oxygen species by human and bacterial pyruvate and 2-oxoglutarate dehydrogenase multienzyme complexes reconstituted from recombinant components. *Free Radic. Biol. Med.* **89**, 642–650 (2015).

60. Nagy, B. et al. Structure of the dihydrolipoamide succinyltransferase (E2) component of the human alpha-ketoglutarate dehydrogenase complex (hKGDHc) revealed by cryo-EM and cross-linking mass spectrometry: implications for the overall hKGDHc structure. *Biochim. Biophys. Acta* **1865**, 129889 (2021).

61. Wu, S. et al. Diverse roles of assembly factors revealed by structures of late nuclear pre-60S ribosomes. *Nature* **534**, 133–137 (2016).

62. Takeda, H. et al. Mitochondrial sorting and assembly machinery operates by β-barrel switching. *Nature* **590**, 163–169 (2021).

63. Zhang, Y. et al. Structure of the mitochondrial TIM22 complex from yeast. *Cell Res.* **31**, 366–368 (2021).

64. Srinivasan, A., Perik, J. & Lill, R. Crystal structures of nucleotide-free and glutathione-bound mitochondrial ABC transporter Atm1. *Science* **343**, 1137–1140 (2014).

65. Diederichs, K. et al. Structural insight into mitochondrial β-barrel outer membrane protein biogenesis. *Nat. Commun.* **11**, 3290 (2020).

66. Lamb, D. C. et al. A second FMN binding site in yeast NADPH-cytochrome P450 reductase suggests a mechanism of electron transfer by diflavic reductases. *Structure* **14**, 51–61 (2006).

67. van Strien, J. et al. CEDAR, an online resource for the reporting and exploration of complexome profiling data. *Biochim. Biophys. Acta Bioenerg.* **1862**, 148411 (2021).

68. Kushnirov, V. V. Rapid and reliable protein extraction from yeast. *Yeast* **16**, 857–860 (2000).

69. Ellenrieder, L. et al. Dual role of mitochondrial porin in metabolite transport across the outer membrane and protein transfer to inner membrane. *Mol. Cell* **73**, 1056–1057 (2019).

70. Araiso, Y. et al. Structure of the mitochondrial import gate reveals distinct preprotein paths. *Nature* **575**, 395–401 (2019).

71. Tucker, K. & Park, E. Cryo-EM structure of the mitochondrial protein-import channel TOM complex at near-atomic resolution. *Nat. Struct. Mol. Biol.* **26**, 1158–1166 (2019).

72. Psakhye, I. & Jentsch, S. Identification of substrates of protein-group SUMOylation. *Methods Mol. Biol.* **1475**, 219–231 (2016).

73. Haan, C. & Behrmann, I. A cost effective non-commercial ECL-solution for Western blot detections yielding strong signals and low background. *J. Immunol. Methods* **318**, 11–19 (2007).

74. Richter-Dennerlein, R. et al. DNAJC19, a mitochondrial cochaperone associated with cardiomyopathy, forms a complex with prohibitins to regulate cardiolipin remodeling. *Cell Metab.* **20**, 158–171 (2014).

75. Thornton, N. et al. Two modular forms of the mitochondrial sorting and assembly machinery are involved in biogenesis of α-helical outer membrane proteins. *J. Mol. Biol.* **396**, 540–549 (2010).

76. Yamano, K., Tanaka-Yamano, S. & Endo, T. Mdm10 as a dynamic constituent of the TOB/SAM complex directs coordinated assembly of Tom40. *EMBO Rep.* **11**, 187–193 (2010).

77. Klein, A. et al. Characterization of the insertase for β-barrel proteins of the outer mitochondrial membrane. *J. Cell Biol.* **199**, 599–611 (2012).

78. Qiu, J. et al. Coupling of mitochondrial import and export translocases by receptor-mediated supercomplex formation. *Cell* **154**, 596–608 (2013).

79. Wang, Q. et al. Structural insight into the SAM-mediated assembly of the mitochondrial TOM core complex. *Science* **373**, 1377–1381 (2021).

80. Ellenrieder, L. et al. Separating mitochondrial protein assembly and endoplasmic reticulum tethering by selective coupling of Mdm10. *Nat. Commun.* **7**, 13021 (2016).

81. Wiedemann, N. et al. Machinery for protein sorting and assembly in the mitochondrial outer membrane. *Nature* **424**, 565–571 (2003).

82. Waizenegger, T. et al. Tob38, a novel essential component in the biogenesis of β-barrel proteins of mitochondria. *EMBO Rep.* **5**, 704–709 (2004).

83. Qi, L. et al. Cryo-EM structure of the human mitochondrial translocase TIM22 complex. *Cell Res.* **31**, 369–372 (2021).

84. Gebert, N. et al. Dual function of Sdh3 in the respiratory chain and TIM22 protein translocase of the mitochondrial inner membrane. *Mol. Cell* **44**, 811–818 (2011).

85. Stiller, et al. Mitochondrial OXA translocase plays a major role in biogenesis of inner membrane proteins. *Cell Metab.* **23**, 901–908 (2016).

86. Smith, P. M., Fox, J. L. & Winge, D. R. Biogenesis of the cytochrome bc_1 complex and role of assembly factors. *Biochim. Biophys. Acta* **1817**, 872–882 (2012).

87. Ott, M., Amunts, A. & Brown, A. Organization and regulation of mitochondrial protein synthesis. *Annu. Rev. Biochem.* **85**, 77–101 (2016).

88. Mick, D. U. et al. Shy1 couples Cox1 translational regulation to cytochrome *c* oxidase assembly. *EMBO J.* **26**, 4347–4358 (2007).

89. Mick, D. U. et al. Coa3 and Cox14 are essential for negative feedback regulation of COX1 translation in mitochondria. *J. Cell Biol.* **191**, 141–154 (2010).

90. Barrientos, A., Korr, D. & Tzagoloff, A. Shy1p is necessary for full expression of mitochondrial COX1 in the yeast model of Leigh's syndrome. *EMBO J.* **21**, 43–52 (2002).

91. Hettema, E. H., Valdez-Taubas, J. & Pelham, H. R. B. Bsd2 binds the ubiquitin ligase Rsp5 and mediates the ubiquitination of transmembrane proteins. *EMBO J.* **23**, 1279–1288 (2004).

92. Gupta, R. et al. Ubiquitination screen using protein microarrays for comprehensive identification of Rsp5 substrates in yeast. *Mol. Syst. Biol.* **3**, 116 (2007).

93. Lin, C. H., MacGurn, J. A., Chu, T., Stefan, C. J. & Emr, S. D. Arrestin-related ubiquitin-ligase adaptors regulate endocytosis and protein turnover at the cell surface. *Cell* **135**, 714–725 (2008).

94. Bykov, Y. S., Rapaport, D., Herrmann, J. M. & Schuldiner, M. Cytosolic events in the biogenesis of mitochondrial proteins. *Trends Biochem. Sci* **45**, 650–667 (2020).

95. Song, J., Herrmann, J. M. & Becker, T. Quality control of the mitochondrial proteome. *Nat. Rev. Mol. Cell Biol.* **22**, 54–70 (2021).

96. Bragoszewski, P., Turek, M. & Chacinska, A. Control of mitochondrial biogenesis and function by the ubiquitin proteasome system. *Open Biol.* **7**, 170007 (2017).

97. Vazquez-Calvo, C., Suhm, T., Büttner, S. & Ott, M. The basic machineries for mitochondrial protein quality control. *Mitochondrion* **50**, 121–131 (2020).

98. Youle, R. J. Mitochondria-striking a balance between host and endosymbiont. *Science* **365**, eaaw9855 (2019).

99. Bömer, U. et al. The sorting route of cytochrome b_2 branches from the general mitochondrial import pathway at the preprotein translocase of the inner membrane. *J. Biol. Chem.* **272**, 30439–30446 (1997).

100. Bender, T., Lewrenz, I., Franken, S., Baitzel, C. & Voos, W. Mitochondrial enzymes are protected from stress-induced aggregation by mitochondrial chaperones and the Pim1/LON protease. *Mol. Biol. Cell* **22**, 541–554 (2011).

101. Böttinger, L. et al. Mitochondrial heat shock protein (Hsp) 70 and Hsp10 cooperate in the formation of Hsp60 complexes. *J. Biol. Chem.* **290**, 11611–11622 (2015).

102. Gerbeth, C. et al. Glucose-induced regulation of protein import receptor Tom22 by cytosolic and mitochondria-bound kinases. *Cell Metab.* **18**, 578–587 (2013).

Acknowledgements

We thank J. Jordan for help with bioinformatics; M. Glickman for providing the His-ubiquitin plasmid; and O. Neuber for discussions. Work included in this study has also been performed in partial fulfilment of the requirements for the doctoral theses of A.G., J.E. and S.M. (University of Bonn) and S.P.S. (University of Freiburg). This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) SFB 1381 project ID 403222702 (to U.S., N.W. and B.F.); PF 202/9-1 project ID 394024777 (to N.P.); FA 332/15-1 project ID 439189341, FA 332/17-1 project ID 446245862 and TRR 152 project ID 239283807 (to B.F.); BE 4679/2-2 project ID 269424439 and SFB 1218 project ID 269925409 (to T.B.); WI 4506/1-1 project ID 406757425 (to N.W.); Germany’s Excellence Strategy (CIBSS EXC-2189 project ID 390939984, to U.S., H.R., N.W., N.P. and B.F.); and the European Research Council (ERC) Consolidator grant no. 648235 (to N.W.).

Author information

Author notes

1. Christoph Mårtensson

   Present address: MTIP, Basel, Switzerland

2. Lars Ellenrieder

   Present address: Novartis, Basel, Switzerland

3. Sebastian P. Straub

   Present address: Sanofi-Aventis (Suisse), Vernier, Switzerland

4. These authors contributed equally: Uwe Schulte, Fabian den Brave, Alexander Haupt, Arushi Gupta

Authors and Affiliations

1. Institute of Physiology, Faculty of Medicine, University of Freiburg, Freiburg, Germany

   Uwe Schulte, Alexander Haupt, Catrin S. Müller, Wolfgang Bildl & Bernd Fakler

2. CIBSS Centre for Integrative Biological Signalling Studies, University of Freiburg, Freiburg, Germany

   Uwe Schulte, Heike Rampelt, Nils Wiedemann, Nikolaus Pfanner & Bernd Fakler

3. Institute of Biochemistry and Molecular Biology, Faculty of Medicine, University of Bonn, Bonn, Germany

   Fabian den Brave, Arushi Gupta, Jiyao Song, Jeannine Engelke, Swadha Mishra & Thomas Becker

4. Institute of Biochemistry and Molecular Biology, ZBMZ, Faculty of Medicine, University of Freiburg, Freiburg, Germany

   Arushi Gupta, Jiyao Song, Christoph Mårtensson, Lars Ellenrieder, Chantal Priesnitz, Sebastian P. Straub, Kim Nguyen Doan, Bogusz Kulawiak, Heike Rampelt, Nils Wiedemann & Nikolaus Pfanner

5. Faculty of Biology, University of Freiburg, Freiburg, Germany

   Sebastian P. Straub

6. Laboratory of Intracellular Ion Channels, Nencki Institute of Experimental Biology, Polish Academy of Sciences, Warsaw, Poland

Bogusz Kulawiak

7. BIOSS Centre for Biological Signalling Studies, University of Freiburg, Freiburg, Germany

Nils Wiedemann & Nikolaus Pfanner

8. Center for Basics in NeuroModulation, Freiburg, Germany

Bernd Fakler

Contributions

U.S., H.R., N.W., N.P., B.F. and T.B. conceived and supervised the project. U.S., T.B., C.S.M., W.B., A.H. and B.F. acquired and evaluated data related to csBN–MS analyses and set-up MitCOM for the CEDAR platform. F.d.B., A.G., J.S., J.E., S.M., C.M., L.E., C.P., S.P.S., K.N.D., B.K., H.R., N.W., N.P. and T.B. performed and analysed experiments related to the analysis of protein assemblies in yeast. U.S., F.d.B., A.H., S.P.S., H.R., B.F. and T.B. prepared figures. B.F., U.S., T.B. and N.P. wrote the paper with the support of all of the authors.

Corresponding authors

Correspondence to Nikolaus Pfanner or Bernd Fakler.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Jens Nielsen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Experimental workflow for determination and evaluation of MitCOM.

Scheme illustrating major materials, tools and intermediates used to generate the MitCOM dataset and to identify protein complexes together with their subunit composition (for details see Methods). Novel or significantly improved tools or procedures are highlighted in red. Quality and usability of MitCOM benefitted particularly from the high-resolution sampling of the blue native gel, the advanced processing of MS-data for determination of accurate protein abundance–mass profiles and the new viewer toolbox including multi-Gaussian component fitting and correlation analysis. Note that the final assignment of protein profile peaks to a distinct complex requires biochemical verification by affinity purification and functional analysis. We used elements of Figs. 1a, 5h, Extended Data Figs. 3b, 5 and an adapted image of Thermo Scientific™ Q Exactive™ HF-X hybrid quadrupole-Orbitrap™ MS for illustration (permission by copyright owner: Thermo Fisher Scientific Inc., Waltham, MA, USA).

Extended Data Fig. 2 Benchmarking of cryo-slicing blue native mass spectrometry.

a, Diagram summarizing number and abundance of proteins identified in the mitochondrial preparation from yeast. We used a mild procedure for mitochondrial preparation such that possible interaction partners from other

cellular compartments remained associated with the mitochondrial surface/outer membrane complexes **b**, Stepwise processing of MS-data as detailed in Materials and Methods using protein Por2 (VDAC2) as an example; the data points and line in black at step 3 represent the ‘abundance over slice number’ profile for Por2. **c**, Profiles (around the abundance peak) of the indicated mitochondrial proteins illustrating equally high resolution over the entire blue native gel range. **d**, Conversion of slices to apparent molecular mass using fit of a sigmoidal function (curve in red) to the masses of the following proteins: 1, Oxoglutarate Dehydrogenase/Ketoglutarate Dehydrogenase complex (3020 kDa); 2, pre-60S ribosome large subunit (2680 kDa); 3, Dimer of F_1F_O-ATP synthase (1250 kDa); 4, III_2IV_2 supercomplex (1000 kDa); 5, III_2IV_1 supercomplex (750 kDa); 6, Monomer of F_1F_O-ATP synthase (600 kDa); 7, SAM-Mdm10 complex (185.5 kDa); 8, TIM22 complex (174 kDa); 9, ATM1 (155 kDa); 10, SAM_{core} complex (Sam50, Sam37, Sam35, 129.4 kDa); 11, succinate dehydrogenase (130 kDa); 12, NADPH-cytochrome P450 reductase (Ncp1, 77 kDa). **e**, Multi-peak fitting by an automated algorithm approximating the profile data obtained for Aim24 (altered inheritance rate of mitochondria) (red line) by a sum of eight Gaussian functions (see Methods). The procedure provided the Gaussian parameters midpoint (apparent molecular mass), amplitude (abundance) and half-width (focusing or variation in size) for the peaks detected in a given profile. Black line represents the result of the multi-component fit, individual Gaussians are colour-coded.

Extended Data Fig. 3 Using the Profile Viewer for interactive data exploration.

Illustration of a selected set of functions (highlighted in red) for targeted data inspection and analysis implemented in the MitCOM Profile Viewer (https://www3.cmbi.umcn.nl/cedar/browse/experiments/CRX36 or, with extended functions, at https://www.complexomics.org/datasets/mitcom). The features include profile visualization and inspection of quantification details, as well as custom-defined MitCOM-wide Pearson correlation analysis for searching constituents of protein complexes. **a**, adjustable Gaussian smoothing of displayed abundance–mass profiles; **b**, inspection of peptide details used for protein quantification (i.e. underlying the selected

datapoint) and generation of the abundance–mass profile(s); **c**, correlation analysis tool. A region of interest is targeted ('target peak') within the profile of a selected protein (left). Correlation analysis is started by clicking the "correlation" tab that opens a range-selection panel in which left and right boundaries of the target peak can be set by dragging the range-slider handles. Simultaneously, a list of 20 proteins exhibiting the highest Pearson correlation with the selected target peak segment is displayed (right, calculation in real time); **d**, normalization of selected sets of proteins to their maximum. A help page with detailed user instructions is available on the Viewer's main window.

Extended Data Fig. 4 Unsupervised matching analysis of MitCOM protein profile segments.

a, Histogram of similarity scores obtained for the segments of the abundance–mass profiles of MitCOM proteins (total of 1.2×10^9 values). The scores integrate (i) correlation of the profile segments of any MitCOM protein with those from all MitCOM proteins (local Pearson correlation, Extended Data Fig. 3c), (ii) ratio of their molecular abundance and (iii) distance of their maxima (see Methods); a value of 1.0 indicates perfect agreement of all three parameters. Pairs in groups of protein profile peaks shown in the figures and/or selected for biochemical verification experiments had score values between 1.0 − 0.88 (highlighted in red). **b**, Pairs of protein profile segments (indicated by the slice numbers given in brackets) exhibiting best similarity scores ('simscore') of the proteins highlighted and/or used for further experiments in the indicated figures.

Extended Data Fig. 5 Visualization of the matching analysis within the MitCOM dataset.

Center, Plot showing t-distributed stochastic neighbourhood distribution (t-SNE) of the entirety of protein abundance profile segments (49,111) represented by dots. Results of the automated Gaussian-fitting (Fig. 1c) were used to define 'seeds' (black dots) for Pearson correlation with the corresponding sections of all 906 MitCOM protein profiles (see also Viewer tool in Extended Data Fig. 3c). Similarity scores (Methods and Extended

Data Fig. 4) were calculated for all protein sections (components) with r > 0.95 (grey dots) and used as distance measure for t-SNE. As a result, close co-localization on this map indicates potential association of the respective proteins into one or more complexes with a defined apparent molecular mass (i.e. localization in Fig. 1c). Top insets are stepwise zooms into the plot resolving a closely co-localized group of protein profile segments representing the main peak of succinate dehydrogenase (respiratory chain complex II) assembled from four subunits Sdh1–4. Note the exquisite co-localization and discrimination of these subunit even at highest magnification factors (indicated in each inset). Bottom insets are magnification of the framed windows in the central plot demonstrating close co-localization of profile segments of biochemically verified subunits of the indicated multi-protein complexes: SAM complex (left), prohibitin/*m*-AAA supercomplex (middle), assembly of Om14 with TOM (right).

Extended Data Fig. 6 Complexity of mitochondrial protein assemblies.

a, Complexity (number of peaks per protein) of mitochondrial assemblies identified in the abundance–mass profiles of the 818 mitochondrial proteins accessible to automated Gaussian fitting. The mean value of the complexity was 6.4 peaks/assemblies per protein (SEM: 0.1, n = 818 mitochondrial proteins). **b**, Mean complexity determined for proteins with the indicated number of predicted transmembrane domains (TM, left), protein mass (middle) and overall abundance in mitochondria (right). **c**, Complexity and size related to submitochondrial localization. Bar graphs summarizing the complexity (left panel, number of peaks per protein) or the apparent molecular size (right panel, mass range bins) of assemblies/complexes identified in the abundance–mass profiles of proteins located in the outer and inner membranes (OM, IM), the intermembrane space (IMS) or the matrix of mitochondria. The plots represent results of 562 proteins for which subcompartmental localization was extracted from SGD GO terms. Relative numbers (left) and abundances (right) of the proteins in the subcompartments are indicated by pie charts. The apparent mass range was binned into equal log-value intervals (bins of 0.25) of the apparent molecular mass. The highest number of peaks (complexity) was observed

for inner membrane proteins. The molecular mass ranges of protein assemblies strongly differ between the mitochondrial subcompartments. Inner membrane complexes are present in high abundance in the mass range of ~500–1,800 kDa, consistent with the presence of oxidative phosphorylation supercomplexes and further large membrane protein complexes[8,74], whereas outer membrane complexes, including protein translocases and metabolite channels, and many matrix complexes display the highest abundance between 100–320 kDa. **d**, Complexity related to functional categories. Mean complexity of different functional categories and subcategories of mitochondrial proteins. Protein count, number of proteins representing these categories in the complexome determined in this work. Regarding the molecular mass distribution of protein assemblies, the two most abundant functional categories, oxidative phosphorylation and metabolism, display major differences (Fig. 2). Proteins involved in oxidative phosphorylation show the highest abundance in a molecular mass range of ~500–1,800 kDa. A fraction of proteins operating in metabolic processes migrate in the high molecular mass range >1,800 kDa, the majority of metabolism-linked proteins, however, migrate below 320 kDa. The category protein biogenesis and turnover shows the highest abundance between 180 and 320 kDa in agreement with the sizes observed for major protein translocases[62,70,71]. Proteins involved in regulatory or redox processes are predominantly found in the low molecular mass range. The category morphology and dynamics displays a broad distribution with a maximum at an apparent mass of >3,200 kDa (Fig. 2) that includes the mitochondrial contact site and cristae organizing system (MICOS) and associated assemblies[1,10].

Extended Data Fig. 7 Localization, complexity and function of non-mitochondrial proteins.

a, Bar graph summarizing the complexity (manually verified number of peaks per protein) for the 660 non-mitochondrial proteins accessible to profile evaluation as in Fig. 1c. **b**, Distribution of subcellular localizations annotated for the 985 non-mitochondrial proteins identified by MS-analysis in the mitochondrial preparation used. Note that the majority of these proteins localized either to cytosol or to ER / Golgi. **c**, Mean complexity determined for the proteins in the indicated subcellular compartments. **d**,

Functional categorization of non-mitochondrial proteins according to their annotation in Uni-Prot/Swissprot and SGD. Functional categories from Fig. 2 were used when applicable. **e**, Bar graph summarizing the complexity determined for the proteins of the indicated functional categories.

Extended Data Fig. 8 Characterization of dynamic protein machineries by MitCOM.

a, Analysis of the SAM complex illustrates the application of MitCOM in the characterization of dynamic protein machinery. The figure shows normalized abundance–mass profiles of SAM subunits, Mdm10 and Tom5. Distinct forms of SAM and TOM complexes are indicated. The SAM complex, also termed TOB complex for topogenesis of β-barrel proteins, inserts the precursors of β-barrel proteins into the mitochondrial outer membrane. SAM consists of several forms that are in exchange with each other and couple to different partner proteins[62,75,76,77,78,79]. The SAM_{core} complex consists of Sam50, Sam37 and Sam35[62,65,77]. Coupling to TOM (transient TOM-SAM supercomplex) facilitates the transfer of precursor proteins from TOM to SAM, whereas the coupling to Mdm10 (SAM-Mdm10 complex) promotes precursor release[62,76,78,80]. In addition, Sam37 plays a dynamic role[62] and its release from SAM_{core} leads to a Sam35-Sam50 subcomplex[81,82]. The apparent overlap of the Sam35-Sam50 subcomplex with Mdm10 in the low molecular mass range does not reflect co-assembly as the peaks can be distinguished by the differences in peak maxima and half-widths. MitCOM efficiently resolves the distinct SAM forms: Sam35-Sam50, SAM_{core}, SAM-Mdm10, and TOM-SAM supercomplexes. **b**, **c**, The interconnected system of the TIM22 complex, the small TIM chaperones Tim9-Tim10, and the SDH complex illustrates a high precision of MitCOM in separation and analysis of complexes of different abundance. The figures show abundance–mass profiles of carrier pathway TIM proteins and SDH (inset in (c): grey box at enlarged scale). The TIM22 complex[63,83] migrates in a main peak of ~200 kDa in MitCOM. For Tim9 and Tim10, two populations are resolved, one bound to the TIM22 complex and the other one representing the soluble hetero-hexameric chaperone complex of the intermembrane space. Sdh3 is found in the SDH complex and the ~20-times less abundant TIM22 complex (c-

normalized abundance vs. b-absolute abundance)[6,63,84]. Note that the precursor of Sdh3 is imported and inserted into the inner membrane via the presequence translocase TIM23 and the OXA translocase, not via the TIM22 complex; mature Sdh3 subsequently assembles into the SDH complex and the TIM22 complex[85]. Sdh4 that is not part of the TIM22 complex does not display a peak at the TIM22 complex (c, insert). MitCOM thus efficiently and quantitatively separates TIM22, Tim9-Tim10 and SDH. **d**, MitCOM quantitatively separates distinct complexes of the F_1F_O-ATP synthase. The figure shows full-range abundance profiles of the subunits of F_1F_O-ATP synthase. The major peak at 1,200 kDa represents the fully assembled dimeric ATP synthase predominating under the mild solubilization conditions used (see Methods), followed by the monomer (at 560 kDa) and the F_1-module (at 250 kDa). Left inset, discrimination of two monomeric ATP synthase complex populations. The abundant peak at 560 kDa lacks the dimerizing subunits Atp20 and Atp21. The primed monomeric form at 590 kDa includes Atp20 and Atp21[49]; note the shoulder-like appearance of the profile peaks of the common subunits caused by the overlap of these two assemblies. Right inset, close co-migration of Atp1 and Atp2 in the low molecular mass range. **e**, Normalized abundance–mass profiles of prohibitins Phb1 and Phb2, of the *m*-AAA protease subunits Yta10 and Yta12, and of the inner membrane-bound J-protein Mdj2. The prohibitin 1/2 (Phb1/2) complex forms an oligomeric scaffold and associates with the protease complex *m*-AAA (Yta10/12)[74]. In mammalian mitochondria, the J-protein DNAJC19 interacts with the prohibitin complex[74], yet it has been unknown if and which yeast mitochondrial J-protein may associate with the prohibitin complex. The inner membrane-bound J-protein Mdj2 migrates in multiple peaks, the largest and most abundant one perfectly co-migrating with the large prohibitin/*m*-AAA supercomplex of ~2.2 MDa. **f**, Affinity purification with tagged Phb1 demonstrates that Mdj2 is the yeast J-protein partner of the prohibitin/*m*-AAA supercomplex. Wild-type (WT) and $Phb1_{ProtA}$ mitochondria were subjected to affinity purification using Protein A sepharose. Bound proteins were eluted with TEV protease and analysed by immunodetection with the indicated antisera. Load 6%, elution 100%. Control proteins: Atp4, subunit of the F_1F_O-ATP synthase; Tim23, subunit of the presequence translocase of the inner membrane. **g**, Normalized

abundance–mass profiles of structural subunits of respiratory complex IV (Cox1, Cox2, Cox4, Cox6) and complex III (cytochrome bc_1 complex; subunit Qcr2), and of the assembly factors Shy1, Coa1 and Mss51. Each of the assembly factors Coa1, Mss51 and Shy1 migrates in multiple peaks in line with their reported presence in various assembly intermediates[86,87,88,89]. A major peak of Shy1 tightly co-migrates with the fully assembled III_2IV_2 respiratory supercomplex of 1 MDa (arrow), in contrast to Coa1 and Mss51. **h**, WT and $Cox4_{His}$ mitochondria were subjected to affinity purification via Ni-NTA agarose and analysed by immunodetection with the indicated antisera. Load 2%, elution 100%. Control proteins: Aco1, aconitase; Hsp60, heat shock protein of 60 kDa. Tagged Cox4 that is not part of early Mss51-containing assembly intermediates[42] co-purifies Shy1 together with the structural subunit Cox2. Whereas qualitative blue native studies suggested that only a small fraction of Shy1 is associated with the full III_2IV_2 supercomplex[88,89,90], MitCOM provides a direct quantitative analysis and demonstrates that a main peak of Shy1 is present at the fully assembled supercomplex, different from other assembly factors such as Mss51 and Coa1.

Extended Data Fig. 9 Characterization of Pth2, Ubp16 and Rsp5.

a, Absolute copy number per yeast cell of the indicated proteins. Values are based on the quantification of mitochondrial proteins grown on different carbon sources by Morgenstern et al.[6]. **b**, Wild-type (WT) and $Tom40_{HA}$ cell extracts were lysed with digitonin and subjected to affinity purification. Proteins were analysed by SDS-PAGE and immunodetection with the indicated antisera. Load 1%, elution 100%. **c**, WT and $Pth2_{His}$ mitochondria were lysed with digitonin and subjected to affinity purification via Ni-NTA agarose. Proteins were analysed by SDS-PAGE and immunodetection with the indicated antisera. Load 0.2% (Tom40, Tom22, Tom70) or 2% (Pth2, Msp1), elution 100%. Control protein: Msp1, mitochondrial sorting of proteins 1. The precursor of Pth2 is imported from the cytosol via the receptor Tom70 and inserted into the outer membrane by the mitochondrial import complex MIM[21]. **d**, Serial dilutions of the indicated strains were

spotted onto full medium with either glucose or glycerol as carbon source and grown at 30 °C. **e**, Serial dilutions of the indicated strains were spotted onto full medium containing glucose or glycerol as carbon source and grown at 24 °C. **f**, Rsp5 preferentially binds to substrates and adaptor proteins via interaction with a PPxY/LPxY motif[31,91,92,93]. Tom70 contains a PPxY motif (aa 71–74) and may thus be involved in recruiting Rsp5 to mitochondria. WT, $Tom40_{HA}$ and $Tom40_{HA}$ *tom70*Δ cell extracts were lysed with digitonin and subjected to affinity purification. Proteins were analysed by SDS-PAGE and immunodetection with the indicated antisera. Load 0.2%, elution 100%. The interaction of Rsp5 with the TOM complex is impaired in the absence of Tom70. **g**, WT and *tom70*Δ cells expressing b_2-DHFR and His-tagged ubiquitin as indicated were lysed under denaturing conditions and subjected to affinity purification via Ni-NTA agarose. Proteins were analysed by SDS-PAGE and immunodetection with antisera against the DHFR domain (upper panel) or ubiquitin (lower panel). b_2-DHFR-Ub, ubiquitin modified b_2-DHFR. Load 0.2%, elution 100%. Precursor ubiquitylation is diminished in the absence of Tom70. **h**, We found an LPxY motif in the DHFR domain of b_2-DHFR. WT yeast cells expressing the standard precursor cytochrome b_2-DHFR^{LPxY} or the mutant form b_2-DHFR^{GGxY} and His-tagged ubiquitin as indicated were lysed under denaturing conditions and subjected to affinity purification via Ni-NTA. Proteins were analysed by SDS-PAGE and immunodetection with the indicated antisera. b_2-DHFR-Ub, ubiquitin modified b_2-DHFR. Input 0.2%, elution 100%. Ubiquitylation of the mutant precursor was reduced. These data suggest that Tom70 is involved in recruiting Rsp5 to the mitochondrial import site. Rsp5 may use Tom70 as an adapter for interacting with precursor proteins and, in case a precursor substrate contains a PPxY/LPxY motif, also engage the motif on the precursor.

Extended Data Fig. 10 Characterization of yeast cells and mitochondria lacking Ubp16, Pth2 or Ubx2.

a, The indicated protein amounts of WT and *ubp16*Δ mitochondria were analysed by SDS-PAGE and immunodetection with the indicated antisera. **b**, WT and *ubp16*Δ mitochondria were lysed with digitonin and analysed by

blue native electrophoresis and immunodetection using an antiserum against Tom22, Tom40 and Hsp60. The levels of TOM, Ubx2 and further tested proteins (a) were not altered in *ubp16*Δ mitochondria, excluding that the lack of Ubp16 led to indirect effects on the protein levels of import or mitoTAD components. **c**, WT and the indicated strains expressing cytochrome b_2Δ-DHFR and His-tagged ubiquitin as indicated were lysed under denaturing conditions and subjected to affinity purification via Ni-NTA. Proteins were analysed by SDS-PAGE and immunodetection with the indicated antisera. b_2Δ-DHFR-Ub, ubiquitin modified b_2Δ-DHFR. Input 0.2%, elution 100%. **d**, Serial dilutions of the indicated strains were spotted onto full medium with either glucose or glycerol as carbon source and grown at 30 °C. **e**, Cell extracts of wild-type (WT), *pth2*Δ, *ubx2*Δ and *pth2*Δ *ubx2*Δ strains were analysed by immunodetection with the indicated antisera. p, precursor; m, mature form of Mdj1. The single deletions of Pth2 and Ubx2 did not lead to an accumulation of the Mdj1 precursor when the complete mitochondrial import machinery including Pam17 was present (see also Fig. 3d)[17]. However, the Mdj1 precursor moderately accumulated in *pth2*Δ *ubx2*Δ double mutants, supporting the view that a fraction of preproteins accumulate even when the mitochondrial import system is fully active[94,95,96,97,98,99,100,101,98]. Thus, at least one of the pathways, Ubx2 or Pth2, has to be active for removal of accumulated precursor proteins. **f**, The indicated protein amounts of WT, *pth2*Δ, *ubx2*Δ and *pth2*Δ *ubx2*Δ mitochondria were analysed by SDS-PAGE and immunodetection with the indicated antisera. **g**, The indicated protein amounts of WT, *pth2*Δ, *pam17*Δ and *pth2*Δ *pam17*Δ mitochondria were analysed by SDS-PAGE and immunodetection with the indicated antisera. **h**, WT, *pth2*Δ, *pam17*Δ and *pth2*Δ *pam17*Δ mitochondria were lysed with digitonin and analysed by blue native electrophoresis and immunodetection with Tom40 specific antisera. The steady state levels of Ubx2, Cdc48 and TOM are not substantially altered in *pth2*Δ and *pth2*Δ *pam17*Δ mutant mitochondria, excluding that the accumulation and ubiquitylation of precursor proteins (Fig. 3d, e) is indirectly caused by a loss of mitoTAD or TOM components. Yeast strains, plasmids and antisera are listed in Supplementary Tables 4–6 (refs. [99,100,101,102]).

Supplementary information

Supplementary Information

Supplementary Fig. 1 and Supplementary Tables 4–6.

Reporting Summary

Peer Review File

Supplementary Table 1

MS data related to identification/quantification of mitochondrial proteins. Sheet 1: MitCOM proteins (total of 906 proteins); sheet 2: mitochondrial proteins with single-peptide detection (total of 49 proteins).

Supplementary Table 2

Data underlying abundance–mass profiles of MitCOM proteins.

Supplementary Table 3

MS data related to identification/quantification of non-mitochondrial proteins (total of 985 proteins).

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain

permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Schulte, U., den Brave, F., Haupt, A. *et al.* Mitochondrial complexome reveals quality-control pathways of protein import. *Nature* **614**, 153–159 (2023). https://doi.org/10.1038/s41586-022-05641-w

- Received: 16 May 2022
- Accepted: 09 December 2022
- Published: 25 January 2023
- Issue Date: 02 February 2023
- DOI: https://doi.org/10.1038/s41586-022-05641-w

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

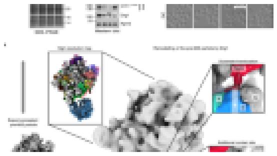
Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

A census of complexes formed by mitochondrial proteins

Nature Research Briefing 25 Jan 2023

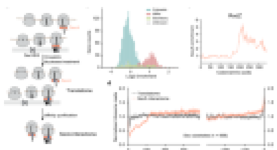
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05641-w



Visualizing maturation factor extraction from the nascent ribosome by the AAA-ATPase Drg1

12 September 2022

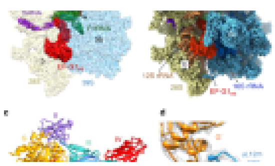
Michael Prattes, Irina Grishkovskaya, … Helmut Bergler



Ribosome profiling reveals multiple roles of SecA in cotranslational protein export

13 June 2022

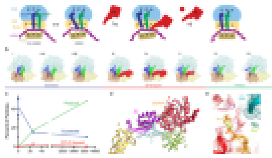
Zikun Zhu, Shuai Wang & Shu-ou Shan



Structures of the human mitochondrial ribosome bound to EF-G1 reveal distinct features of mitochondrial translation elongation

31 July 2020

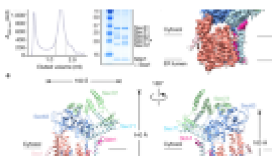
Ravi Kiran Koripella, Manjuli R. Sharma, … Rajendra K. Agrawal



[Time-resolved cryo-EM visualizes ribosomal translocation with EF-G and GTP]

13 December 2021

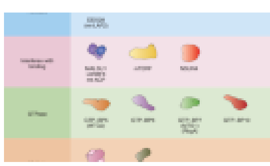
Christine E. Carbone, Anna B. Loveland, … Andrei A. Korostelev



[Structure of the post-translational protein translocation machinery of the ER membrane]

31 December 2018

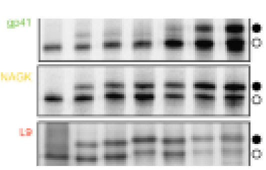
Xudong Wu, Cerrone Cabanos & Tom A. Rapoport



[Mitoribosome assembly comes into view]

11 August 2021

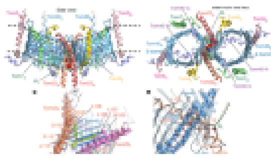
Marie Sissler & Yaser Hashem



[Transmembrane but not soluble helices fold inside the ribosome tunnel]

07 December 2018

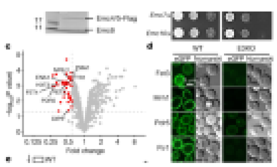
Manuel Bañó-Polo, Carlos Baeza-Delgado, … Ismael Mingarro



[Structure of the mitochondrial import gate reveals distinct preprotein paths]

10 October 2019

Yuhei Araiso, Akihisa Tsutsumi, … Toshiya Endo



[Structure of the ER membrane complex, a transmembrane-domain insertase]

03 June 2020

Lin Bai, Qinglong You, … Huilin Li

- Article
- Open Access
- [Published: 25 January 2023]

Visualization of translation and protein biogenesis at the ER membrane

- Max Gemmer,
- Marten L. Chaillet,
- Joyce van Loenhout,
- Rodrigo Cuevas Arenas,
- Dimitrios Vismpas,
- Mariska Gröllers-Mulderij,
- Fujiet A. Koh,
- Pascal Albanese,
- Richard A. Scheltema,
- Stuart C. Howes,
- Abhay Kotecha,
- Juliette Fedry &
- …
- Friedrich Förster

Show authors

Nature **volume 614**, pages 160–167 (2023)

- 20k Accesses
- 241 Altmetric
- Metrics details

Abstract

The dynamic ribosome–translocon complex, which resides at the endoplasmic reticulum (ER) membrane, produces a major fraction of the human proteome[1,2]. It governs the synthesis, translocation, membrane

insertion, *N*-glycosylation, folding and disulfide-bond formation of nascent proteins. Although individual components of this machinery have been studied at high resolution in isolation[3,4,5,6,7], insights into their interplay in the native membrane remain limited. Here we use cryo-electron tomography, extensive classification and molecular modelling to capture snapshots of mRNA translation and protein maturation at the ER membrane at molecular resolution. We identify a highly abundant classical pre-translocation intermediate with eukaryotic elongation factor 1a (eEF1a) in an extended conformation, suggesting that eEF1a may remain associated with the ribosome after GTP hydrolysis during proofreading. At the ER membrane, distinct polysomes bind to different ER translocons specialized in the synthesis of proteins with signal peptides or multipass transmembrane proteins with the translocon-associated protein complex (TRAP) present in both. The near-complete atomic model of the most abundant ER translocon variant comprising the protein-conducting channel SEC61, TRAP and the oligosaccharyltransferase complex A (OSTA) reveals specific interactions of TRAP with other translocon components. We observe stoichiometric and sub-stoichiometric cofactors associated with OSTA, which are likely to include protein isomerases. In sum, we visualize ER-bound polysomes with their coordinated downstream machinery.

Main

In mammalian cells, the vast majority of membrane proteins, secreted proteins and soluble proteins of most organelles are synthesized at the ER membrane. A cleavable N-terminal signal peptide emerging from the ribosome targets most secretory pathway proteins to the ER[1,2], where the nascent chain elongation is continued, concomitant with its translocation across or insertion into the ER membrane. During its elongation cycle, the ribosome recruits aminoacyl-tRNAs (aa-tRNAs) matching the mRNA codons in the aminoacyl (A)-site, forms the peptide bond between the amino acid and the nascent chain, and translocates the mRNA–tRNA moiety[3]. The GTP-dependent eEF1a and eEF2 support the required tRNA movements and motions of the ribosomal small subunit (SSU) with respect to the large subunit (LSU).

Ribosomes bind to the dynamic ER translocon complex[2]. Its invariant core module—the heterotrimeric protein-conducting channel SEC61—faces the ribosomal exit tunnel. To facilitate protein transport and to accommodate the signal peptide, SEC61 can switch from closed to open conformations[8,9]. SEC61 associates with distinct cofactors that reflect the requirements of different substrates. The translocon-associated protein complex (TRAP), a hetero-tetrameric transmembrane protein complex supporting the insertion of many signal peptides[10,11], is a near-stoichiometric ER translocon component[12]. Low-resolution studies revealed interactions of TRAP with ribosomal RNA (rRNA) expansion segments and ribosomal subunit protein 38e[13] (RPL38e), but the molecular details remain unresolved given the absence of an atomic model of TRAP. OSTA, which is responsible for co-translational *N*-glycosylation of substrates, is observed in at least 50% of translocon particles in mammalian cells[14]. Although the structure of OSTA and its specific association with the ribosome and SEC61 have been studied extensively[13], its native interactions, including those with biogenesis cofactors such as ER chaperones remain unknown. In addition to the SEC61–TRAP and SEC61–TRAP–OSTA translocons, a ribosome-bound translocon specialized in the insertion of multipass transmembrane proteins, has recently been isolated and analysed structurally[15,16]. Here we have used cryo-electron tomography (cryo-ET) to visualize the elongating ribosome at the ER membrane and its downstream translocation and biogenesis machinery.

Subtomogram analysis of ribosome complexes

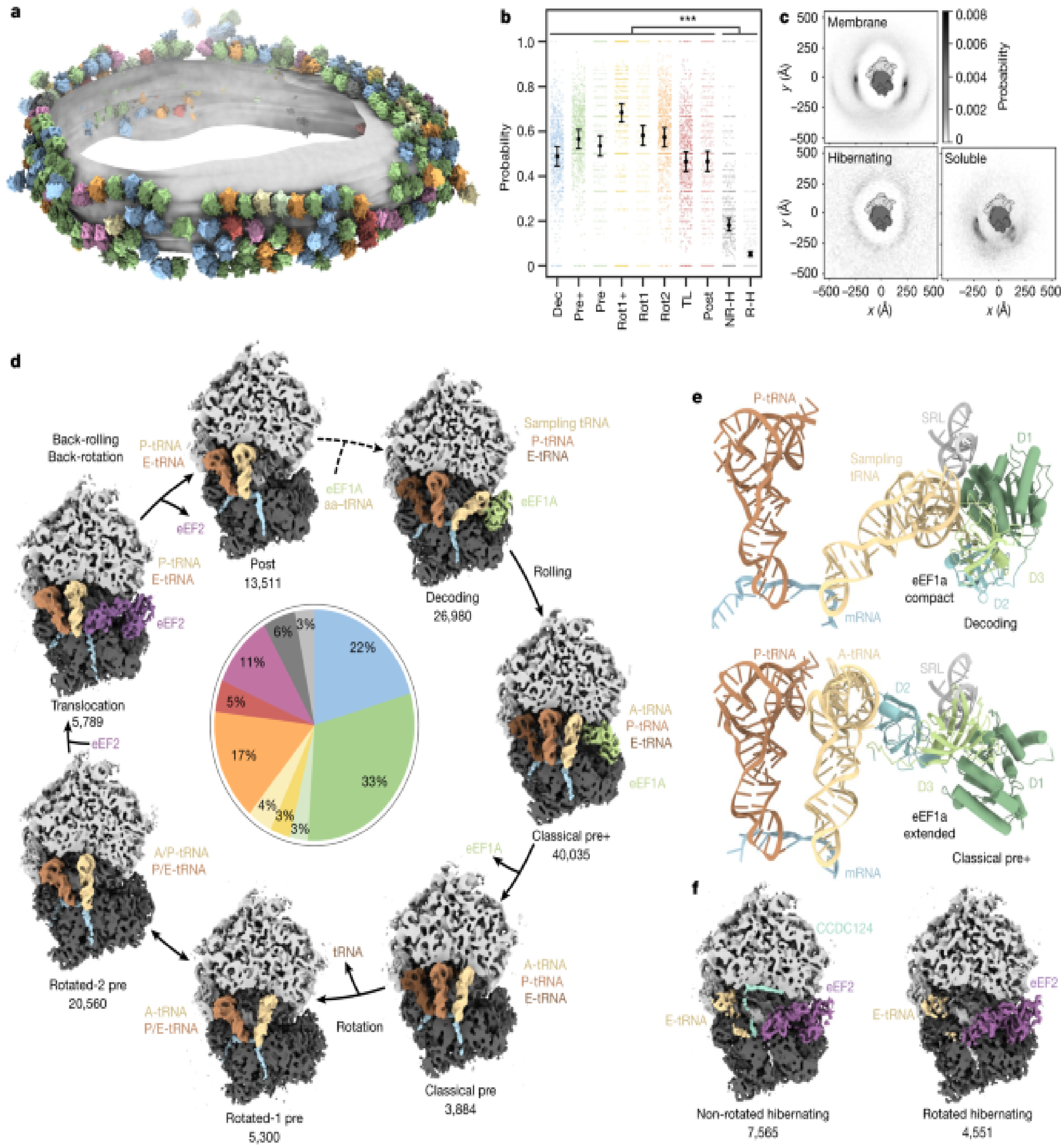
To analyse the elongation cycle of ER-bound ribosomes and the associated ER translocon complex we rapidly (within about 1 h) isolated ER-derived vesicles (microsomes) from HEK 293F cells for subsequent cryo-ET imaging (Extended Data Fig. 1 and Supplementary Fig. 1). We acquired a large dataset (869 tilt series) of frozen hydrated vesicles and used a regularized single-particle analysis approach to analyse the membrane-associated ribosome particles[17] (Extended Data Fig. 1). Extensive subtomogram analysis reveals the most abundant ribosome and translocon states. Altogether, we distinguish ten ribosomal intermediate states and four translocon variants, as well as two translocon-bound chaperones at

resolutions ranging from 4 to 10 Å, which allows the identification of ribosomal intermediate states on the basis of high-resolution structures of isolates (Extended Data Fig. 2 and Supplementary Table 1).

Ribosomal intermediates and 3D distribution

We first dissected the translational states of the ribosome pool consisting of membrane-bound and residual soluble particles. Focusing on the orientation of the SSU and association of tRNAs and elongation factors, we classified the particles into ten distinct states (Extended Data Figs. 1c and 2). To assess their translational activity, we examined the relative 3D distribution of the particles from the classes using a reciprocal neighbourhood probability analysis, which is indicative of integration into polysomes (Fig. 1a–c and Extended Data Fig. 3). Particles from eight classes (89%) show probability hotspots proximal to the ribosomal mRNA entrance and exit (E)-sites characteristic for membrane-bound and cytosolic ribosomes and consistent with previous lower-resolution analyses[18,19]. By contrast, two classes show a featureless neighbour distribution, implying that these particles are not involved in polysomes. The reconstructions of these two classes do not have tRNA bound in the peptidyl (P)-site and resemble known hibernating ribosome complexes bound to eEF2[20].

Fig. 1: Captured human ribosomal states and spatial distribution.



a, Different ribosome states mapped back onto one exemplary ER-derived vesicle ($n = 869$ tomograms from one experiment, two independent replicates; Extended Data Fig. 4). **b**, Probabilities of ribosome states being present in polysomes. Black circles show the modelled mean and error bars represent the 95% confidence interval (CI) ($n = 132{,}371$ ribosomes with the 869 tomograms included as a random effect). Hochberg-adjusted *P* values were determined using a two-sided Wald-test. *P* values for comparison between hibernating and elongating states were all smaller than 2×10^{-16}. The small scattered points represent the frequencies of events per

tomogram. TL, translocation; NR-H, non-rotated hibernating; R-H, rotated hibernating. **c**, Neighbour distribution of ER membrane-bound, hibernating and soluble ribosome particles. The membrane resides in the plane of the image. **d**, Observed active intermediates positioned in a model of the human elongation cycle. All reconstructions were filtered to 7 Å resolution. The ribosome is clipped for visualization. The tRNAs are colour-coded with respect to a complete cycle. The abundance of each state is indicated in the pie chart, colour-coded as in **a**. **e**, Close-up views of ribosome-bound compact eEF1a in the decoding-sampling state (PDB: 4CXG) and in the classical pre+ state (4C0S). D1 to D3 indicate the extended eEF1a domains 1 to 3. **f**, Reconstructions of two distinct ribosome states lacking tRNA at the P-site (hibernating states).

To assess the physiological relevance of our preparation, we analysed the distribution of ribosomal intermediate states in situ using focused ion beam (FIB) milled human cells. Although the lower yield of this approach resulted in substantially fewer particles (5,818) and reduced classification depth, it confirmed the high abundance of factor-bound classes (around 70%), and their presence on polysomes (Extended Data Fig. 4). The approximately 66% factor-bound ribosome complexes ex vivo exceed the abundance in previous ribosomal purification from HEK cells involving size-exclusion chromatography (around 8% in ref. [21]) (Extended Data Fig. 4f). Consistent with this previous cryo-electron microscopy (cryo-EM) analysis[21] and the high abundance of eEF1a and eEF2 in proteomics data of the sample (Supplementary Fig. 2), we identify eEF1a and eEF2 as ribosome-binding factors. Nevertheless, we stress that lysis and the isolation conditions may affect intermediate complexes and their abundances, which may eventually be overcome when higher resolution is achievable for FIB cryo-ET studies of human cells.

Elongation cycle intermediates

To further analyse the polysome-associated ribosomal classes, we attempted to position them in the context of the elongation cycle as modelled on the basis of knowledge from previous in vitro reconstitution studies[3,21,22,23,24] (Fig. 1d). Although one class could not be conclusively assigned functionally (Supplementary Fig. 3), the remaining seven states are

consistent with previous structural or biochemical data. The elongation cycle model commences with delivery of aa-tRNAs to the ribosome by GTP-bound eEF1a (Fig. 1d, decoding state). Approximately 22% of ribosomes in our data adopt an unrotated state, with clear densities for the tRNAs in the P- and E-sites and the eEF1a–tRNA ternary complex, which we assigned to a decoding population[25] (Fig. 1e). The position of eEF1a in our decoding map differs slightly from a previously reported decoding state in polysomes purified from HEK cells[21], which may be a result of differences in the preparation protocols. The position of eEF1A in our decoding complex rather resembles a codon sampling state obtained by inhibiting eEF1a GTP hydrolysis[25] (Extended Data Fig. 5a,b). We speculate that the decoding population observed in our data may be explained by ribosomes testing non-cognate tRNAs that do not trigger GTP hydrolysis and occur more frequently than cognate tRNAs in the cell.

Next, we observe a highly abundant intermediate (33%) that has not been described previously: whereas the tRNA is accommodated in the canonical A-site and the SSU 'rolls' into the classical pre-configuration, eEF1a is bound to the ribosome in an extended conformation, which matches crystal structures of purified eEF1A–GDP[26] and its bacterial homologue EF-Tu–GDP[27] (Fig. 1e, Extended Data Fig. 5c,d and Supplementary Video 2). To analyse the pre+ state at higher resolution, we rapidly isolated soluble ribosomes and imaged them with cryo-EM single-particle analysis (SPA). Approximately 30% of particles were in the pre+ state, yielding a focused reconstruction of eEF1a with specific side chains of domain 3 (approximately 3.5 Å resolution) unambiguously identifying eEF1a (Extended Data Fig. 6a–g). In the classical pre+ state, eEF1a domains 1 and 3 interact with the sarcin–ricin loop (SRL) of the 28S rRNA (Extended Data Fig. 6h,i), whereas domain 2 blocks the A/T site and contacts the A-site tRNA. In a model of the human elongation cycle, we propose that the classical pre+ state may follow the decoding state, in which eEF1a still adopts a compact conformation. Although we cannot rule out that other factors observed at this site in situ could have been displaced by eEF1A (Extended Data Fig. 4) during the purification, the occurrence of the eEF1A bound classical pre+ state in purified samples indicates the possibility that eEF1a may remain bound to the ribosome during conformational switching to the extended form. This observation is different from bacteria, where no

factors are observed in situ on the abundant pre-like A/P state[28] and suggests differences in eukaryote post-hydrolysis proofreading, possibly involving eEF1A[29,30,31]. The functional relevance of a possible eEF1a-bound classical pre+ state remains to be further investigated with complementary methods.

Next, we observe a previously described classical pre state, which we propose to occur after eEF1a fully dissociates from the ribosome, as the SSU and tRNAs remain unchanged (3%). We then identified two rotated states in our data: the rotated-1 pre state resulting from dissociation of a tRNA (4%), and the much more highly populated rotated-2 pre state with the tRNAs in hybrid A/P and P/E positions (17%). In contrast to previous studies of cytosolic polysomes[21], we found 5% of ribosomes in a state resembling a translocation intermediate associated with eEF2 and tRNAs in the canonical P- and E-sites. GTP hydrolysis seems to have occurred as indicated by the disordered switch I loop (Supplementary Fig. 4). This state resembles the late translocation intermediate post-3 state[24], which would be consistent with kinetic studies in the bacterial system[32]. Finally, we observe a similar state with tRNAs in the P- and E-sites and without eEF2, which is in good agreement with the post translocation (post) state (Fig. 1d).

Finally, we note that the assigned positions of the three most abundant states we observe are consistent with the elongation rate-limiting steps: decoding and pre+ correspond to proofreading steps, whereas rotated-2 precedes translocation.

Hibernating ribosomes and ER stress

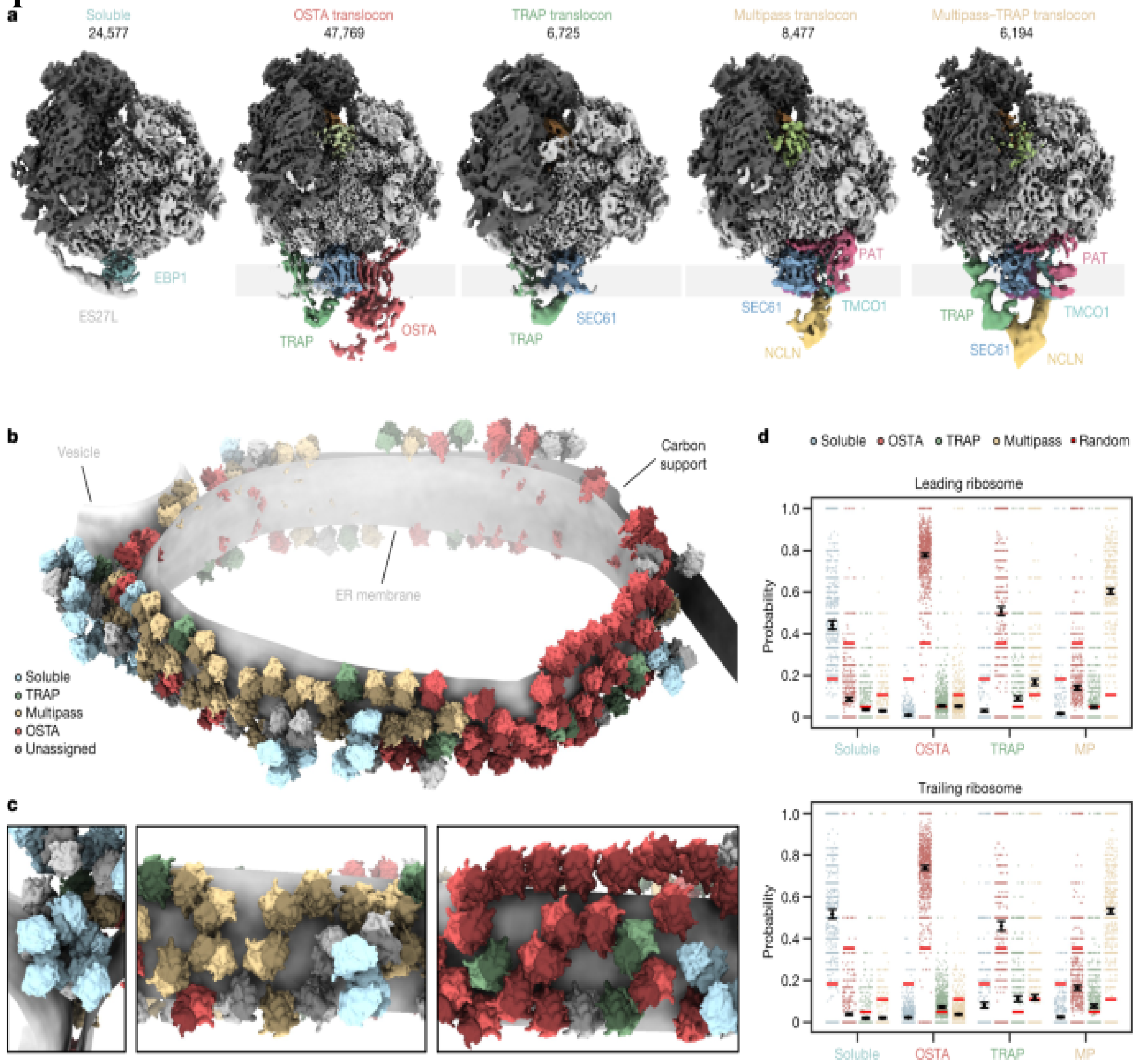
Membrane-bound hibernating ribosomes group into two major populations (Fig. 1f). A non-rotated state with a tRNA bound at the exit (E)-site and the protein CCDC124 occupying the P-site (7%) differs from a similar structure of the cytosolic hibernating ribosome[20] by eEF2 binding. We also detected a second rotated ribosome state (5%), which features eEF2 and from which CCDC124 is absent, analogous to the cytosolic hibernating ribosome[20]. To investigate the physiological role of hibernating ribosomes, we also imaged microsomes from dithiothreitol (DTT)-treated HEK cells, in which elongation activity should be reduced[33]. Notably, we observed almost

exclusively hibernating ribosomes upon treatment with DTT (Supplementary Fig. 5a,b). Thus, the abundance of ER-bound hibernating ribosomes depends strongly on cell state, and possibly also on cell density, as observed for cytosolic hibernating ribosomes[20]. We cannot rule out induction of some hibernating ribosomes by lysis, which must be considered when interpreting the relative abundances.

Native ER translocon distribution

We then grouped the particles according to their structural features near the ribosomal exit tunnel into five different classes (Fig. 2a): one soluble ribosome class and four membrane-bound ribosome classes. Approximately 30% of particles, mostly ‘top views’, were not assigned to any of these distinct five classes owing to insufficient signal or the missing wedge (Extended Data Fig. 1a). Soluble ribosomes are associated with EBP1 embraced by expansion segment 27L (ES27L) at the exit tunnel[34] (Supplementary Fig. 6), whereas membrane-bound ribosomes (64,208 particles) contact four distinct ER translocon complexes (Fig. 2a): the most populated SEC61–OSTA–TRAP (69% of ER-bound particles) and SEC61–TRAP translocons (10%) have previously been identified in cryo-ET datasets of dog pancreatic ER-derived microsomes[14]. The ER translocons in the remaining two classes (21%) have a common larger component, with one of them also harbouring TRAP. The common density has been observed but not identified previously in ER microsomes from HEK 293T cells upon knockout of OSTA subunit STT3a[7]. This translocon component resembles a recently discovered transmembrane protein complex responsible for insertion of multipass transmembrane proteins[15]. In addition to SEC61, the multipass translocon comprises the insertase TMCO1, the PAT complex and the nicalin–TMEM147–NOMO complex[16,35]. To confirm the assignment of our density to the multipass (TMCO1–PAT–nicalin–TMEM147–NOMO) translocon, we knocked out CCDC47, a component of the PAT complex. Indeed, cryo-ET data of the ΔCCDC47 microsomes did not display the density at the position of the protein in the isolated multipass translocon[15] (Supplementary Fig. 7a,b). Thus, the major translocon types in wild-type HEK ER microsomes are SEC61-multipass, SEC61-multipass–TRAP, SEC61–OSTA–TRAP and SEC61–TRAP (Fig. 2a).

Fig. 2: Organization of soluble and ER membrane-associated ribosome populations.



a, Locally filtered reconstructions of different soluble and ER membrane-associated ribosome populations resulting from 3D classification focused near the exit tunnel. **b**, Segmented representation of one tomogram of an ER-derived vesicle (*n* = 869 tomograms from 1 experiment). Populations from **a** are mapped back into the reconstruction and coloured accordingly. **c**, Close-up views of the segmentation from **b**. **d**, The probability of encountering ER-associated ribosomes from **a** as leading or trailing neighbour. Black circles show the modelled mean and error bars represent the 95% CI (*n* = 45,751 ribosomes with the 869 tomograms included as a random effect). The small, scattered points represent the frequencies of

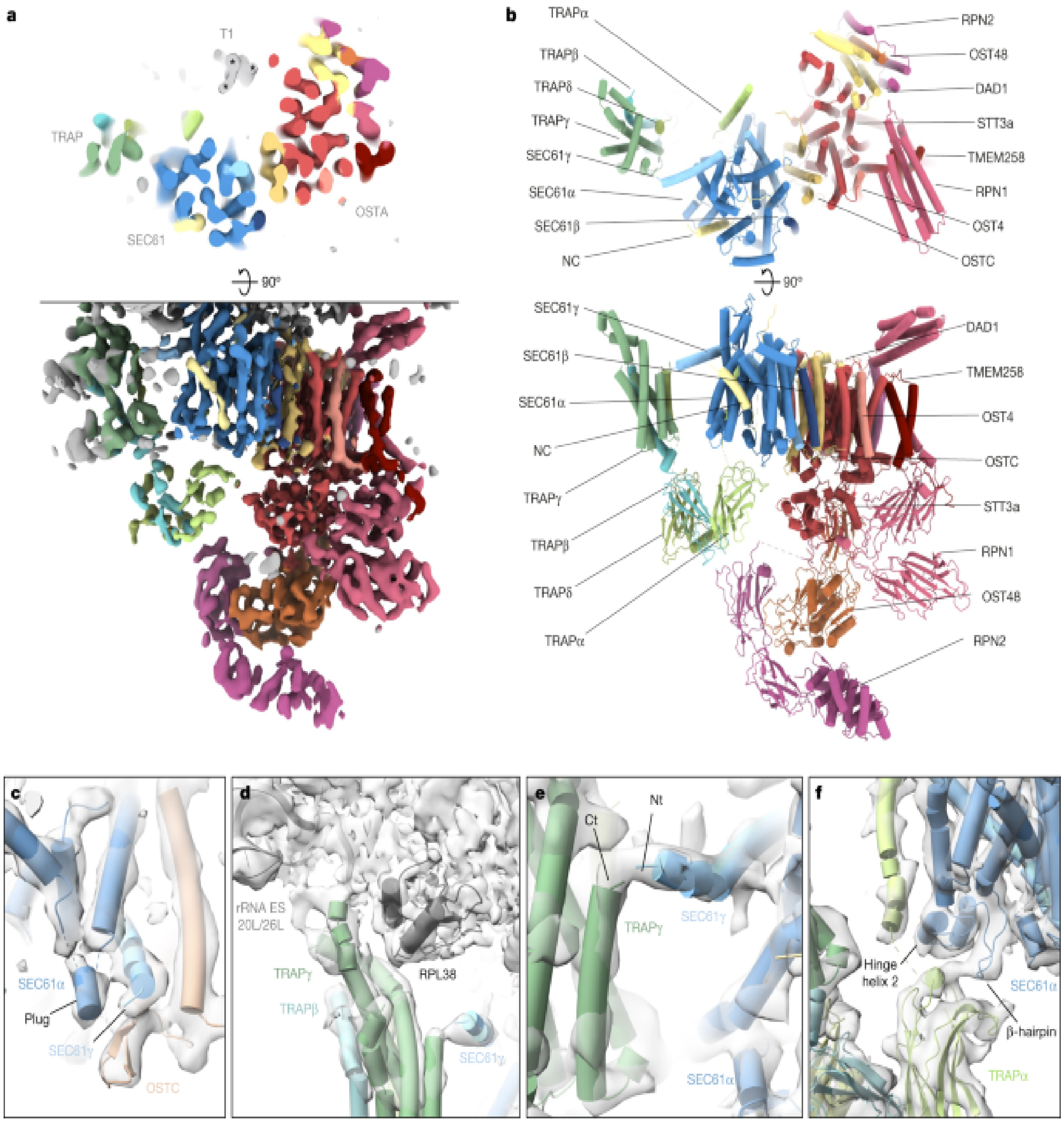
events per tomogram. The random association probability (bright red lines) is the overall abundance of the ribosome populations. MP, multipass.

Mapping back the particles of these different ribosome–translocon populations in the original tomograms indicate clustering according to their translocon type (Fig. 2c). To further examine their polysomal organization, we used our neighbour probability analysis in the context of leading and trailing ribosome neighbours, which reflect late and early stages of translation, respectively (Fig. 2d and Extended Data Fig. 7). This statistical approach indicates a strong segregation of ribosomes bound to OSTA-containing and multipass translocons, as well as soluble EBP1 (Fig. 2d). SEC61–TRAP translocons have less tendency to pair among themselves. They also neighbour OSTA-containing and multipass translocons, where they are preferably found as a trailing polysome neighbour (Extended Data Fig. 7c,d). Thus, nascent peptides preferentially encounter SEC61–TRAP translocons early in their biogenesis. Later, the membrane-bound translocon machineries specialize—this is consistent with recent studies on different model substrates[36].

Architecture of the SEC61–OSTA–TRAP translocon

Ribosome-centred refinement of the most abundant population, the SEC61–OSTA–TRAP translocon, yielded a 4.2 Å-resolution structure (focused on the LSU) with poorly resolved transmembrane helices (TMHs) (7–10 Å resolution). Recentring on the ER luminal domains resolved those at improved resolution (6–8 Å) (Fig. 3a,b and Extended Data Fig. 8a–d). A composite of both densities enabled us to build a near-complete atomic model using AlphaFold[37].

Fig. 3: Atomic model of the most abundant ER translocon.



a, Top view (top) and side view (bottom) of the translocon-centred reconstruction of SEC61–TRAP–OSTA. **b**, Atomic model of the ER translocon built from cryo-EM structures (PDB: 3JC2 and 6S7O) and AlphaFold predictions. **c**–**f**, Close-up views showing the molecular model placed into the segmented density maps. **c**, The plug helix of SEC61α contacts the SEC61γ C terminus and the luminal OSTC β-hairpin. SEC61α transmembrane helix 4 (TMH4) and SEC61β were removed for clarity. **d**, The cytosolic TRAPγ domain associates with rRNA expansion segments ES20L and ES26L and the ribosomal protein L35. **e**, The TRAPγ C terminus

contacts the N terminus of SEC61γ. Ct, C terminus; Nt, N terminus. **f**, The luminal TRAPα domain interacts with a β-hairpin of the SEC61α hinge region and the TRAPα transmembrane helix contacts the second helix of the hinge region. SEC61α TM7 to TM10 were removed for clarity.

The SEC61 channel opens its lateral gate to the lipid membrane[8,9] (Fig. 3a,b and Extended Data Fig. 8d). As in previous cryo-ET studies[12], the lateral gate accommodates a pronounced helical density, which matches the position of the signal peptide in isolates[6,38] and may represent an average of the signal peptides of the different proteins synthesized at the ER membrane. Moreover, a density is discernable near the ribosomal exit tunnel that may correspond to an average of nascent chains (Extended Data Fig. 8e). The luminal part of SeC61 reveals a short α-helix, which we assigned to the SEC61α plug (Fig. 3c and Extended Data Fig. 8f,g). This hallmark feature of SEC61 was not resolved in lower-resolution cryo-ET studies[12] and higher-resolution structures of solubilized ribosome–SEC61 complexes[6,7]. Here, we observe the plug in a displaced conformation stabilized by SEC61γ and the oligosaccharyltransferase complex subunit OSTC. This arrangement resembles the yeast post-translocon, in which SEC63 stabilizes the plug[39].

To investigate the structural deviations of SEC61–TRAP–OSTA when bound to hibernating ribosomes, we reconstructed the inactive SEC61–TRAP–OSTA from the DTT-stressed microsomes. Although the SEC61 plug closes in the inactive complex, the density reveals an open lateral gate accommodating a helical density (Supplementary Fig. 5c–f). Since signal peptides can be cleaved co-translationally[40], this helix might correspond to a pool of cleaved signal peptides or to an unknown specific peptide.

An AlphaFold-based model of TRAP could be fitted unambiguously into the SEC61–OSTA–TRAP translocon map, requiring only minor repositioning of single transmembrane helices and removal of low-confidence segments (Extended Data Fig. 9a–c). The assembly model does not display notable clashes and density in the lumen coincides with predicted *N*-glycosylation sites of TRAPα and TRAPβ, further supporting our assignment (Extended Data Fig. 9d).

As previously observed[16], the cytosolic domain of TRAPγ tethers TRAP to RPL38e and the rRNA expansion segments ES20L and ES26L (Fig. 3d). Our results reveal the position of the TRAPα transmembrane helix, separated by a 2–3.5 nm lipid density from the major transmembrane part of TRAP, which comprises TRAPγ, TRAPβ and TRAPδ. In addition, we visualize a contact between the previously unresolved C terminus of TRAPγ and the amphipathic SEC61γ N-terminal helix at the cytosolic face of the membrane (Fig. 3e and Extended Data Fig. 9e). The fibronectin fold domains of TRAPα, TRAPβ and TRAPδ form the luminal part of TRAP, where they may interact with nascent proteins in a confined space. Near the luminal end of its transmembrane helix, the FG and BC loops of the TRAPα fibronectin-like domain associate with the SEC61α hinge region (Fig. 3f), which bridges the pseudo-symmetric N- and C-terminal halves of SEC61. Finally, we observe that TRAP association is not restricted to laterally open SEC61 as the SEC61-multipass–TRAP translocon displays a closed lateral gate both in the presence of CCDC47, as recently shown[41], as well as in the absence of CCDC47 (Supplementary Fig. 7h,i).

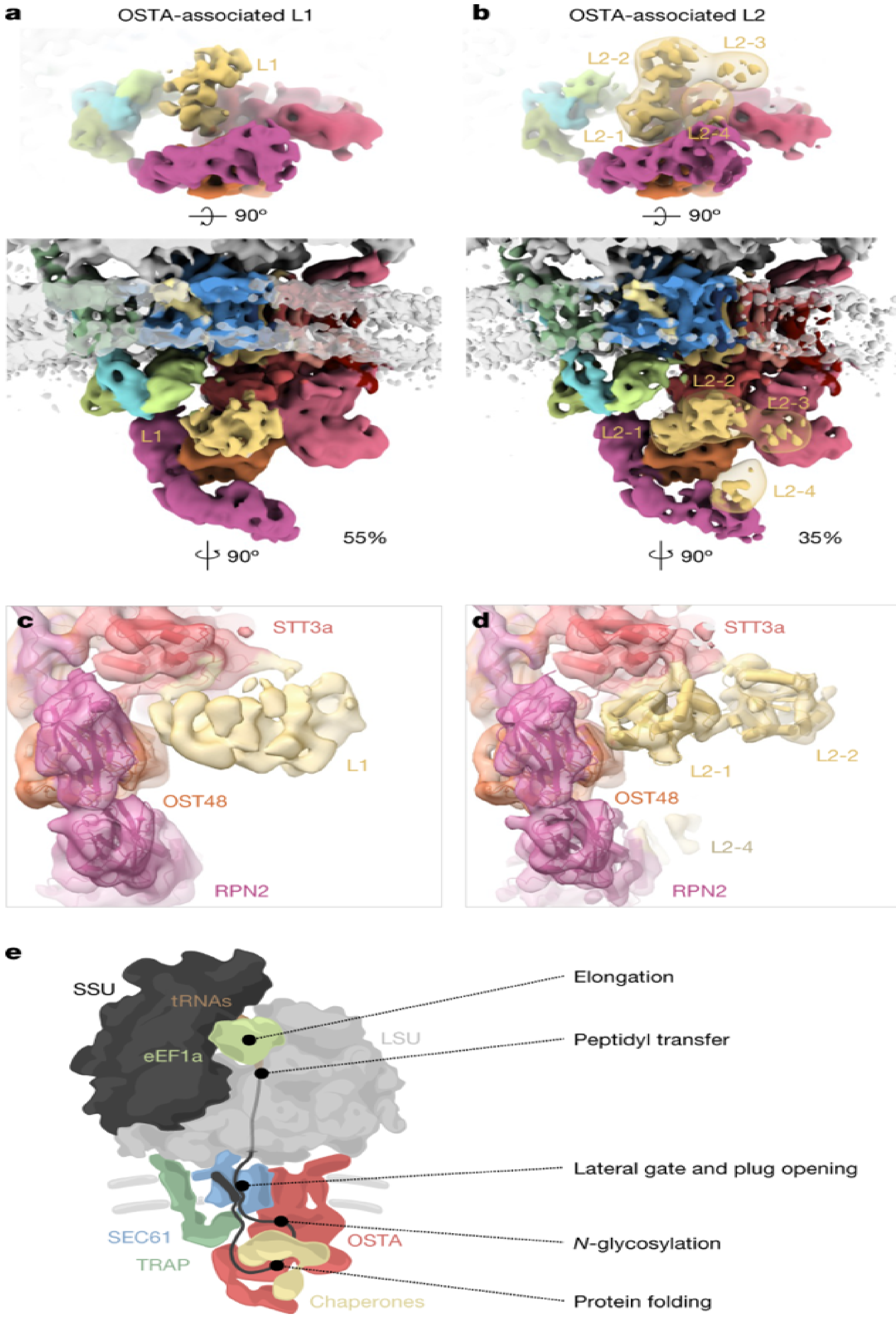
Cellular and biochemical studies indicate that TRAP is required for the biogenesis of proteins that exhibit signal peptides with weak helical propensity due to glycine and proline residues[11]. Preproteins with pronounced hydrophobic helical signal peptides are subject to stronger pulling forces than TRAP-dependent preproteins of, for example, prion protein[42], presumably owing to the lower affinity of their signal peptides for the lateral gate. Although the structure of TRAP–SEC61 does not provide an obvious mechanism of action for the TRAP complex, it enables to formulate a hypothesis. When signal peptides traverse SEC61 head-on and enter the lumen, they contact the luminal TRAPα domain[43]. We speculate that the growing nascent chain pushing against the SEC61 hinge-bound TRAPα domain might then open the SEC61 lateral gate via an allosteric mechanism and expose its hydrophobic surface to accommodate the signal peptide. Alternatively, it was suggested during the revision of this work that lipid bilayer modulation induced by TRAP, which can indeed be observed in our membrane-embedded structure, could promote insertion of signal peptides[44]. Further studies will be required to evaluate the mechanistic function of the TRAP interactions revealed in this study.

Native OSTA and its associated factors

The cryo-ET structure is in excellent agreement with the cryo-EM SPA structure of solubilized OSTA[4], which lacks the RPN2 N-terminal domain (Extended Data Fig. 10a). To complete the atomic model, we fitted the corresponding AlphaFold models into the most membrane-distal part of our map (Extended Data Fig. 10b,c). However, the SEC61–TRAP–OSTA model does not explain a transmembrane helix structure (T1)—approximately 15 kDa in size—comprising three transmembrane helices and a characteristic amphipathic helix facing the cytosol (Fig. 3a and Extended Data Fig. 10d–f). T1 is intercalated between STT3a TMH9 and the C terminus of the TRAPα TMH, resulting in the formation of a lipid-filled cavity near the hinge region of SEC61. In line with OSTA association and cavity formation, T1 was observed only in the OSTA-containing ER translocon (Extended Data Fig. 10g,h). The glucosyltransferases acting upstream of OSTA or the dolichyldiphosphatase I acting downstream of OSTA are candidates for T1[45], but neither atomic model provides an acceptable fit. Thus, further investigation will be required to determine the molecular identity of T1.

We observed weak density associated with the luminal domain of STT3a. For higher-resolution insights into possible sub-stoichiometric binding partners, we performed classification focused on the SEC61-proximal luminal face of the OSTA, which revealed three distinct populations: (1) OSTA without accessory factors (11%), (2) OSTA in complex with a globular density of approximately 35 kDa (L1, 54%), and (3) OSTA in complex with a density of approximately 60 kDa (L2, 35%) (Fig. 4a,b and Extended Data Fig. 10i–l). L1 associates with negatively charged residues at the C terminus of STT3a (amino acids 667–676). L2 comprises four approximately equally sized domains (Fig. 4b), of which domains L2-1 and L2-2 compete for the same binding site with L1. Domain L2-4 binds the N-terminal domain of RPN2, and L2-3 does not interact with OSTA. Whereas L2-1 and L2-2 reveal secondary structure elements, L2-3 and L2-4 bind the flexible RPN2 N-terminal domain and are poorly resolved.

Fig. 4: Co-translational ER biogenesis factors and summary.



a
OSTA-associated L1
L1
90°
L1
55%
90°
b
OSTA-associated L2
L2-2
L2-3
L2-1
L2-4
90°
L2-2
L2-3
L2-1
L2-4
35%
90°
c
STT3a
L1
OST48
RPN2
d
STT3a
L2-1
L2-2
OST48
L2-4
RPN2
e
SSU
tRNAs
LSU
eEF1a
SEC61
TRAP
OSTA
Chaperones
Elongation
Peptidyl transfer
Lateral gate and plug opening
N-glycosylation
Protein folding

a,**b**, Top view (top) and front view (bottom) of the accessory factors L1 (**a**) and L2 (**b**) associated with the SEC61–TRAP–OSTA translocon. The transparent map represents L2 filtered to a resolution of 20 Å. **c**,**d**, Close-up view of the interaction site between STT3a and L1 (**c**) or L2 (**d**). Domains a and b of the L2 candidate protein PDIR were placed into domain L2-1 and L2-2, respectively. **e**, Model of the main protein translation and translocation machinery at the ER membrane.

To our knowledge, L1 and L2 have not previously been observed in OSTA complexes purified from HEK cells[4]; they are likely to represent transiently binding proteins. The ER contains many chaperones that assist in protein biogenesis, which are prime candidates for L1 and L2[46]. Among the ER chaperones, prolyl isomerase cyclophilin B is most abundant in the sample (Supplementary Table 2) and shows the best agreement in shape and size with L1, which is, however, too small and globular for unambiguous assignment. Protein disulfide isomerases[47] (PDIs) show a good fit with the characteristic four-domain structure of the larger L2 well (Fig. 4d and Extended Data Fig. 10m–p). The negative charges at the interacting site of STT3a would be consistent with the interaction pattern observed for PDIs with calnexin and calreticulin[48] (Extended Data Fig. 10q). PDIs are highly abundant in the sample (Supplementary Table 2) with glycoprotein-specific family member PDIA3 (also known as ERp57), probably representing L2[49]. The transient recruitment of an oxidoreductase to OSTA is plausible, as its post-translational counterpart OSTB features a constitutive oxidoreductase[50] (N33 (also known as Tusc3)).

Conclusions

In summary, extensive classification of cryo-ET data visualizes the process of ER-associated translation and the dynamic recruitment protein biogenesis factors in the context of polysomes (Fig. 4c). This study on the ensemble of secretory proteins synthesized in the cell complements biochemical analyses[36,41] and forms the basis for future investigation of the biogenesis of specific proteins and the change of the machinery in distinct cellular states and diseases.

Methods

CRISPR–Cas9 knockout of CCDC47

FreeStyle 293-F cells (Thermo Fisher Scientific, R79007) were transfected with the plasmid pSpCas9(BB)-2A-Puro (PX459) V2.0 from the F. Zhang laboratory (Addgene plasmid 62988) containing the 20-bp single guide RNA (sgRNA) target sequence 5′-CACCGGTACACGGTGAACTCGTGCG-3′, PAM: AGG or 5′-CACCGGGAGGAAGCGGGCGAGGTGC-3′, PAM:GGG. Transfection was performed using Lipofectamine 2000 (Thermo Fisher Scientifiic, 11668019) according to the manufacturer's instructions using 1 μg DNA per ml of culture at a cell density of 1×10^6 cells per ml. Cells were cultured for 48 h in FreeStyle 293 expression medium (Thermo Fisher Scientific, 12338018) on an orbital shaker (120 RPM) at 37 °C and supplemented with 5% CO_2. Two days after transfection, cells were collected and resuspended in complete Dulbecco's Modified Eagle Medium (DMEM) (Thermo Fisher Scientific, 11966025) (supplemented with 10% fetal bovine serum (Thermo Fisher Scientific, 10100147) and GlutaMAX-I (Thermo Fisher Scientific, 35050061)) with 0.5 μg ml^{-1} Puromycin (InvivoGen, ant-pr-1). Subsequently, cells were plated in T175 flasks (Thermo Fisher Scientific, 159910) and grown for 7 days in complete DMEM with 0.5 μg ml^{-1} Puromycin with periodical medium exchange or sub-culturing when confluency was reached.

After 7 days of Puromycin selection, surviving cells were dislodged, collected, and resuspended at 5 cells per ml in conditioned complete DMEM. One-hundred and fifty microlitres per well of cell suspension was plated into sterile 96-well plates and cultured for 14 days. Cell colonies derived from single cells were used for further cell expansion. After 14 days in culture, conditioned complete DMEM was exchanged for FreeStyle medium and cell colonies transferred into 24-well plates. Subsequently, cells were grown to confluency and further expanded into 6-well plates and 10-cm dishes before analysis.

Cell culture

HeLa and U2OS cells (from ATCC, CVCL_0042 and CVCL_0030 in Cellosaurus.org, respectively) were grown in standard tissue culture conditions (37°, 5% CO2) in DMEM Glutamax (Gibco). HEK 293F cells (Thermo Fisher Scientific, R79007) were grown in suspension in FreeStyle medium with 120 rpm agitation. Cell lines were not authenticated and were tested for negative mycoplasma.

ER-vesicle preparation

HEK 293F wild-type or CCDC47 knockout cells ($0.5–1 \times 10^6$ cells per ml, 50 ml) were collected and washed (3 times with PBS, at 300*g*, 5 min, 4 °C). HEK 293F cells used for ER stress studies were treated with 10 mM DTT for 2 h before collection. Cells were resuspended in lysis buffer (2–4 ml, 10 mM HEPES-NaOH pH 7.4, 250 mM sucrose, 2 mM $MgCl_2$, 0.5 mM DTT, protease inhibitor cocktail (Roche)) and lysed using a Isobiotec cell cracker (5–10 passes, 14 µm clearance, on ice). The lysate was cleared (1,500*g*, 2–3 × 5 min, 4 °C, in 2 ml tubes) using a cooled tabletop centrifuge. Vesicles were pelleted (10,000*g*, 10 min, 4 °C), and washed with resuspension buffer (10 mM HEPES, 250 mM sucrose, 1 mM $MgCl_2$, 0.5 mM DTT). The pellet was resuspended at a concentration of ~50 mg ml^{-1} determined by A_{280}, frozen in liquid nitrogen and stored at −80 °C until further use. The supernatant was used for proteomics as control.

Twenty micrograms of microsomes were used for SDS–PAGE followed by immunoblotting using antibodies against SEC61α (Abcam, ab15575; 1:1,000), TRAPγ (Sigma Aldrich, hpa014906; 1:1,000) and CCDC47 (Abcam, ab241608; 1:1,000).

Mass spectrometry data acquisition

Approximately 100 µg of the isolated ER-microsome and cytosolic fraction (supernatant) were digested using an S-Trap micro-MS column (protifi) according to the vendor’s protocol.

Proteins were solubilized in lysis buffer (10% SDS, 100 mM Tris, pH 8), reduced (100 mM TCEP), alkylated (400 mM CAA in isopropanol) and denatured (27.5% phosphoric acid). For protein trapping, samples were

flown over an S-Trap micro spin column, (10,000g, 30 s) and further washed with binding buffer (100 mM triethylammonium bicarbonate (TEAB) buffer, in 90% methanol). Protein digestion was achieved with an overnight incubation at 37 °C using a water bath (Grant Instruments, JB Academy) after the addition of digestion buffer (10% trypsin, 2% lysine, 50 mM Tris). Protein peptides were retrieved by washing with elution buffer (50 mM Tris), using a tabletop centrifuge (10,000*g*, 1 min).

Eluted peptides were lyophilized and dissolved in 2% formic acid prior to liquid chromatography–mass spectrometry (LC–MS/MS) data acquisition. MS data were acquired using an Ultimate 3000 RSLC nano system (Thermo Scientific) coupled to an Exploris 480 (Thermo Scientific). Three technical replicates of each sample were measured. Peptides were first trapped in a pre-column (Dr. Maisch Reprosil C18, 3 µm, 2 cm × 100 µm) prior to separation on the analytical column packed in-house (Poroshell EC-C18, 2.7 µm, 50 cm × 75 µm), both columns were kept at 40 °C in the built-in oven. Trapping was performed for 10 min in solvent A (0.1% v/v formic acid in water), and the elution gradient profile was as follows: 0–10% solvent B (0.1% v/v formic acid in 80% v/v acetonitrile) over 5 min, 13–44% solvent B over 37 min, 44–100% solvent B over 4 min, and finally 100% B for 4 min before re-equilibration in 100% A for 8 min. The mass spectrometer was operated in a data-dependent mode. Full-scan mass spectra were collected in a mass range of *m*/*z* 350–1,300 Thomson (Th) in the Orbitrap at a resolution of 60,000 after accumulation to an AGC target value of 10^6 with a maximum injection time of 50 ms. In-source fragmentation was activated and set to 15 eV. The cycle time for the acquisition of MS/MS fragmentation scans was set to 1 s. Dynamic exclusion properties were set to n = 1 and to an exclusion duration of 10 s. HCD fragmentation (MS/MS) was performed with a fixed normalized collision energy of 27% and the mass spectra acquired in the Orbitrap at a resolution of 30,000 after accumulation to an AGC target value of 10^5 with an isolation window of *m*/*z* = 1.4 Th.

Raw data were processed using the MaxQuant software[51] version 2.0.1.0 with standard settings applied. In brief, the extracted peak lists were searched against the reviewed Human UniProtKB database (date 15 July 2021; 20,353 entries), with an allowed precursor mass deviation of 4.5 ppm and an allowed fragment mass deviation of 20 ppm. Cysteine

carbamidomethylation was set as static modification, and methionine oxidation, N-terminal acetylation as variable modifications (maximum 5 modifications per peptide allowed). Both LFQ quantification and ‘match between runs’ were enabled. The iBAQ values in Supplementary Fig. 4b are approximate absolute abundances of the identified proteins derived by the normalization of the summed peptide intensities by the number of theoretically observable peptides for a given protein.Raw data were processed using the MaxQuant software[51] version 2.0.1.0 with standard settings applied. In brief, the extracted peak lists were searched against the reviewed Human UniProtKB database (date 15 July 2021; 20,353 entries), with an allowed precursor mass deviation of 4.5 ppm and an allowed fragment mass deviation of 20 ppm. Cysteine carbamidomethylation was set as static modification, and methionine oxidation, N-terminal acetylation as variable modifications (maximum five modifications per peptide allowed). Both LFQ quantification and ‘match between runs’ were enabled. The iBAQ values in Supplementary Fig. 4b are approximate absolute abundances of the identified proteins derived by the normalization of the summed peptide intensities by the number of theoretically observable peptides for a given protein.

Grid preparation

ER vesicles were diluted in resuspension buffer to a concentration of 2–3 mg ml^{-1} and 2 μl were applied onto a glow-discharged lacey carbon grid (Quantifoil). Four m,icrolitres of BSA-conjugated gold beads (10 nm, UMC Utrecht) diluted in resuspension buffer without sucrose were added and mixed with the sample on grid. Grids were immediately blotted from the backside for 5–6 s and plunged into a mix of liquid ethane and propane using a manual plunger.

For the adherent cell lines (Hela and U2OS), cells were seeded on R2/2 holey carbon on gold grids (Quantifoil) coated with fibronectin in a Mattek dish and incubated for 24 h. The suspension HEK 293F cells were grown to mid-log phase, and the cells were then directly pipetted onto glow-discharged R2/1 Carbon on Copper grids (Quantifoil). Grids were immediately blotted from the back for 10 s and plunged into liquid ethane propane mix using a manual plunger.

Lamella preparation

Lamellae were prepared using an Aquilos FIB-SEM system (Thermo Fisher Scientific). Grids were sputtered with an initial platinum coat (10 s) followed by a 10 s gas injection system (GIS) to add an extra protective layer of organometallic platinum. Samples were tilted to an angle of 15° to 22° and 12 µm wide lamellae were prepared. The milling process was performed with an ion beam of 30 kV energy in 3 steps : (1) 500 pA, gap 3 µm with expansion joints, (2) 300 pA, gap 1 µm, (3) 100 pA, gap 500 nm. Lamellae were finally polished at 30–50 pA with a gap of 200 nm.

Data acquisition

We acquired 869 tilt series on a Talos Arctica (Thermo Fisher Scientific) operated at an acceleration voltage of 200 kV and equipped with a K2 summit direct electron detector and energy filter (Gatan). Images were recorded in movies of 7–8 frames at a target defocus of 3 µm and an object pixel size of 1.72 Å. Tilt series were acquired in SerialEM (3.8)[52] using a grouped dose-symmetric tilt scheme[53] covering a range of ±54° with an angular increment of 3°. The cumulative dose of a series did not exceed 80 e^{-} Å^{-2}.

Lamella data used in this analysis has been collected in one session on a pool of grids of human cell lines. Twenty-seven tilt series were acquired on six different lamellae on a Talos Arctica (same instrument as above). Images were recorded in movies of 5–8 frames at a target defocus of 4 µm and an object pixel size of 2.17 Å. Tilt series were acquired in SerialEM using a grouped dose-symmetric tilt scheme covering a range of ±60° with a pre tilt of ±10° and an angular increment of 3°. The cumulative dose of a series did not exceed 70 e^{-} Å^{-2}.

Reconstruction and particle localization

Video files of individual projection images were motion-corrected in Warp (1.0.9)[54] and combined into stacks of tilt series with the determined contrast transfer function (CTF) parameters. The combined stacks were aligned using the gold fiducials in IMOD (4.10.25)[55]. Per-tilt CTF estimation for entire tilt

series was performed in Warp and full deconvoluted tomograms were reconstructed by weighted back projection at a pixel size of 20 Å. Ice thickness was determined manually for a subset of 50 tomograms and results in an average thickness of 156 nm. Particle coordinates were determined by template matching against a reconstruction of a human 80S ribosome filtered to 40 Å and downsampled to match the tomogram pixel size (20 Å) using pyTOM (0.994)[56]. Most false-positive hits were manually removed in pyTOM. The determined positions of ribosomes were used to extract subtomograms and their corresponding CTF volumes at a pixel size of 3.45 Å (2× binned) in Warp. Video files of individual projection images were motion-corrected in Warp[54] and combined into stacks of tilt series with the determined CTF parameters. The combined stacks were aligned using the gold fiducials in IMOD[55]. Per-tilt CTF estimation for entire tilt series was performed in Warp and full deconvoluted tomograms were reconstructed by weighted back projection at a pixel size of 20 Å. Ice thickness was determined manually for a subset of 50 tomograms and results in an average thickness of 156 nm. Particle coordinates were determined by template matching against a reconstruction of a human 80S ribosome filtered to 40 Å and downsampled to match the tomogram pixel size (20 Å) using pyTOM[56]. Most false-positive hits were manually removed in pyTOM. The determined positions of ribosomes were used to extract subtomograms and their corresponding CTF volumes at a pixel size of 3.45 Å (2× binned) in Warp.

Lamellae data were processed as above with slight variations. Video files of individual projection images were motion- and CTF-corrected in Warp and combined into stacks of tilt series. The combined stacks were aligned using patch tracking in IMOD. CTF estimation for entire tilt series was performed in Warp and full tomograms were reconstructed by weighted back projection at a pixel size of 17.36 Å. Ice thickness was determined manually and was found to be <200 nm for all lamellae. Particle coordinates were determined by template matching against a reconstruction of a human 80S ribosome filtered to 40 Å using downsampled to match the tomogram pixel size (17.36 Å) pyTOM. The determined positions of ribosomes were used to extract subtomograms and corresponding CTF volumes at a pixel size of 8.68 Å (4× binned) in Warp.

Subtomogram analysis

The extracted subtomograms were aligned in RELION (3.1.1)[57] using a spherical mask with a diameter of 300 Å against a reference of an 80S ribosome obtained from a subset of the same data. The extracted subtomograms were aligned in RELION (3.1.1)[57] using a spherical mask with a diameter of 300 Å against a reference of an 80S ribosome obtained from a subset of the same data. The aligned particles were refined in M (1.0.9)[17] using the reconstructions of the two half maps as a reference and a tight soft mask focused on the LSU at a pixel size of 3.45 Å. Particles were subjected to 2–3 rounds of refining image warp grid, particle poses, stage angles, volume warp grid, defocus and pixel size. After refinements, new subtomograms and their corresponding CTF volumes were extracted at a pixel size of 6.9 Å (4× binned) and subjected to 3D classification (without mask, without reference, $T = 4$ and classes = 50) to sort out remaining false positives, poorly aligned particles, and lone LSUs. The remaining 134,350 particles were used for subsequent focused classification steps to dissect ribosomal intermediate states or translocon variants.

Classification of ER ribosome populations

All 134,350 particles were subjected to 3D classification (without reference, with soft mask, $T = 4$, classes = 20) in RELION, focused on the area at the ribosomal tunnel exit including the membrane and translocon. Particles were sorted into SEC61–TRAP-bound, SEC61–TRAP–OST-bound, SEC61-multipass-bound and EBP1-bound ribosomes and a combined class of ribosomes with ambiguous densities. Ribosomes with ambiguous densities were subjected to two further classification rounds and sorted the respective class from above until no further separation could be achieved. Ribosomes that associated with the EBP1 were designated ‘soluble’, ribosomes associated with translocon variants were designated ‘membrane-bound’ and ribosomes associated with ambiguous densities were designated ‘unidentified’.

Subtomograms of the multipass translocon were recentered by 17 nm from the centre of the ribosome towards SEC61 and extracted in M at a voxel size of 6.9 Å. Subsequently, subtomograms were classified focused on the luminal domains of TRAP and NCLN (with reference of all multipass translocons, with soft mask, $T = 4$, classes = 3) or focused on the cytosolic

domain of CCDC47 (with reference, with mask, $T = 3$, classes = 2). The TRAP-multipass translocon was further refined using local angular searches in RELION or, to obtain ribosome-centred reconstructions of the multipass translocon populations, subtomograms were recentered again by 17 nm towards the centre of the ribosome in M and subjected to another round of refinement.

Refinement of the OST translocon

The 42,215 best-correlating particles (5,554 particles were poorly aligned) of the OST-bound ribosome were used for refinement focused on the LSU in M using the same parameters as above at a pixel size of 1.72 Å (unbinned), which resulted in a reconstruction at an overall resolution of ~4 Å. However, densities of OST or TRAP in the ER lumen were poorly resolved. To improve local resolution of the translocon components, the reconstruction was recentered by 19.5 nm from the centre of the ribosome towards the OST translocon and subtomograms were extracted in M at a pixel size of 3.45 Å. The particles were aligned in RELION using the average of the recentered reconstruction of the OST translocon as reference and a tight soft mask focused on SEC61, TRAP and OST. Subsequently, the aligned particles were refined in M as above at a pixel size of 1.72 Å resulting in a reconstruction at an overall resolution of 8 Å. Local resolutions estimated using M[17] ranged from 6–7 Å for the OST and 8–9 Å for TRAP and the N-terminal domain of RPN2, indicating flexibility. Local refinement focused on the TRAP complex did not improve its resolution, presumably because the protein complex was too small to provide sufficient signal for reliable refinement.

After refinement in M, translocon-centred OST-particles were extracted at a pixel size of 6.9 Å and subjected to classification in RELION (without reference, with mask, $T = 10$, classes = 4) focused on the chaperone binding site. The resulting classes were refined in M as above using masks focusing on SEC61, TRAP, OST and chaperone.

Classification of ribosomal intermediates

Ribosomal intermediate states were obtained by hierarchical classification focused on the rotation of the SSU and on the tRNA and elongation factor

binding sites. First, all 134,350 particles were classified into classes of ribosomes with non-rotated and rotated SSU (with reference, with soft tight mask focused on SSU, $T = 4$, classes = 2). Subsequently, non-rotated and rotated particles were each subjected to two rounds of classification (with reference, with mask focused on tRNA and elongation factor binding site, $T = 10–20$, classes = 10–20). Classes with fragmented densities, such as pre/pre+, rotated−1/rotated−1+, non-rotated idle/translocation, were separated in the second round of classification (with reference, with mask focused on tRNA and elongation factor binding site, $T = 10–20$, classes = 2–4).

Classification of intermediate states was first performed for individual populations of ER translocon-bound or soluble ribosomes, which revealed similar results for each population. However, to improve performance of classification, especially for translocon-associated populations with a low number of particles, we pooled all translocon and soluble populations and performed classification of intermediates on the entire dataset. Subsequently, particle sets of individual intermediate states were dissected according to the translocon-associated and soluble ribosome populations.

The classification workflow was repeated four times to assess the technical uncertainties of 3D classification, which was determined at 5% to 15% and correlates inversely with class size. To assess experimental reproducibility, we combined two smaller datasets of ER-derived vesicles (31 tomograms, 6,101 particles; 58 tomograms, 3,836 particles) with the large dataset (869 tomograms, 134,350 particles) and processed them as described above. After obtaining classes of intermediate states, particle numbers were determined for each dataset and class.

The classification workflow was applied to in situ data with slight variations: extracted subtomograms were used for 3D classification with image alignment against a low pass filtered 80S ribosome map as reference in RELION to exclude false positive. The remaining 5,818 ribosome subtomograms were refined in RELION and re-extracted in Warp at a pixel size of 4.34 Å (2× binned). Two times-binned subtomograms were refined in RELION with a mask on the LSU prior to a first round of 3D classification without image alignment with a mask on the SSU to separate rotated from non-rotated ribosomes. A second round of classification was performed

using a mask positioned on the tRNA and elongation factors sites, optimizing the mask extension and class number to this data in order to yield stable classes despite limited resolution and particle number. The different classes were finally subjected to iterative refinement in M.

Refinement of intermediate states

Classes of ribosomal intermediate states were simultaneously refined in M at a pixel size of 1.72 Å (unbinned) using tight masks focused on the entire 80S ribosome, tRNAs and elongation factors, which were individually generated for each intermediate. Refinement of image warp grid, particle poses, stage angles, volume warp grid, defocus and pixel size were performed iteratively (2–3 iterations). Globally or locally filtered and sharpened maps were generated by M and used for visualization or model building.

Model building

Initial models for each chain of SEC61 and the OST were downloaded from the Alphafold database[58]. A polyalanine helical stretch was manually built to account for the plug density. The OSTA chains were manually docked into the higher-resolution OSTA SPA map EMD-10110, followed by refinement through an iterative cycling between phenix (1.20.1) refine[59], isolde (1.0b5)[60] and Coot (0.9.8.2)[61]. The initial model for TRAP was built using AlphaFold Colab[37] and Coot[61]. The initial model for TRAP was built using AlphaFold Colab for multimeric complexes[62] and was divided into the transmembrane part and the luminal part. Each model was manually fitted into our subtomogram average (STA) density in UCSF Chimera (1.14.0)[63], followed by normal-mode guided refinement using iMODFIT (1.51)[64]. Long flexible loops not visible in our density were manually removed from the models. SEC61, OSTA and luminal TRAP domains were fitted and refined into a STA centred on the OST, while the TRAP transmembrane helices were fitted and refined into the original ribosome-centred STA, in which they were better defined. Each model was refined using iterative cycling between phenix refine, Isolde and Coot. Models were then combined for one last round of refinement together in the OST centred STA. Validation was

performed using Molprobity (4.5.1)[65]. UCSF ChimeraX (1.3.0)[63] was used for visualization of all models and reconstructions.

Single-particle analysis

Suspension HEK 293F cells were grown to mid-log phase (0.5–1 × 10^6 cells per ml, 50 ml). Cells were pelleted at 500 g for 5 min and washed twice in ice cold PBS and resuspended in 10 mM Hepes KOH, pH 7.5, 250 mM sucrose, 2 mM magnesium acetate, 0.5 mM DTT, 0.5 mM PMSF, protease inhibitor tablets). Cells were lysed with 30 passages through a 21-gauge needle. The lysate was cleared by centrifugation steps at 1,000*g* for 10 min, 1,500*g* for 15 min and 20,000*g* for 20 min. The final supernatant was loaded onto a 1 M sucrose cushion and spun at 300,000*g* for 1 h. The final ribosomal pellet was resuspended in lysis buffer and snap frozen in liquid nitrogen. For grid preparation, 3.5 µl of the ribosome preparation was pipetted onto glow-discharged R 3.5/1 2 nm C holey grids (Quantifoil) and blotted for 2.5 s at force 0 using a Vitrobot (Thermo Fisher Scientific) before subsequent plunging into liquid ethane.

Single-particle cryo-EM data were acquired on a Titan Krios (Thermo Fisher Scientific) equipped with a cold FEG, Falcon 4i detector and Selectris X energy filter 10 eV slit at a pixel size of 0.729 Å per pixel. A total of 17,000 movies was acquired with EPU 3 (Thermo Fisher Scientific) in EER format. A cumulative dose of 40 e^{-} $Å^{-2}$ was used.

The data was processed in Relion 3.1.1. Movies were motion-corrected and CTF was estimated. Particles were picked with the logpicker and reconstructed at a pixel size of 6 Å per pixel for subsequent 2D classification, followed by 3D classification with image alignment to exclude false-positive and low-quality particles. A total of 66,000 particles was then subjected to 3D classification without image alignment using a mask on the A tRNA site and the GTPase centre. 19,000 particles were selected in a class corresponding to the classical pre+ state, refined, re-extracted at 1.0 Å per pixel and refined again. CtfRefine was performed followed by another round of refinement. Masks on the A-site tRNA site and elongation factor, as well as on the peptidyl transferase centre were used for

particle subtraction and focused refinements to improve the quality of the maps in these regions.

For model building, a previous crystallographic structure of eEF1A in the extended GDP bound conformation (PDB 4C0S) was used as starting model and was first briefly refined in real space in the higher-resolution crystallographic electron density map using Isolde and phenix refine, in order to improve the starting geometry of the model. The resulting model was then refined in our map through iterative cycling between phenix refine[59], Isolde[60] and Coot[61]. The model was validated using Coot[61] and Molprobity[65].

Sequence conservation

The degree of sequence conservation was determined using the ConSurf server[66] using 150 homologous sequences with a sequence identity ranging from 35%–95%. The conservation score was plotted onto the surface of the respective protein model in UCSF Chimera.

Polysome analysis

For the neighbourhood analysis, ribosome positions and orientations were read from the RELION star files resulting from subtomogram alignment in a python script (Python 3.8.11, Numpy 1.20.3, Scipy 1.7.1). For each ribosome we determined distance vectors between itself and its n closest neighbours ($n = 4$), excluding neighbours further than 100 Å. The vectors were rotated with the inverse orientation of the respective ribosome, resulting in the coordinates of neighbours in the coordinate system of an ER-bound ribosome with the *xy* plane corresponding to the ER membrane. These vectors were sampled on a 3D-histogram with voxels corresponding to 15^3 Å^3 and divided by the total number of analysed neighbours to indicate the probability of finding a neighbouring ribosome particle in each voxel. The plots were projected on the *xy* plane to visualize the density of neighbours surrounding ER-bound and soluble ribosomes.

A threshold was chosen to identify clusters for trailing and leading neighbours. For ER-bound neighbours a binary mask was created in the 3D-

histogram above a probability of $P = 0.0005$, while for soluble ribosomes the threshold was put at $P = 0.0003$. Both masks were dilated by two voxels. The soluble and ER-bound trailing masks were combined in a trailing mask for the whole dataset, and the same procedure was performed for the leading mask. The masks were used to annotate associations of ribosome pairs in a polysome. A trailing–leading connection was confirmed if the neighbour localized in the trailing–leading mask area and the analysed ribosome also positioned in the leading–trailing area of the respective neighbour (that is, the inverse calculation).

The trailing/leading states of neighbours were used in R to fit a multinomial mixed-effects logistic regression model (mclogit 0.9.4.2[67] in R 3.6.1). The ribosome's state was used to predict probabilities of leading and trailing states, where the tomogram index was used as a random effect to account for sample and imaging variation. We used the same model to predict probabilities of translation states in polysome chains. For visualization, the probabilities were extracted with their 95% confidence interval, representing the region of 95% certainty that the modelled mean is the population mean. Variation between tomograms was shown by calculating the frequency of certain events per tomogram—for example, the 42nd tomogram might have 7 pre+ ribosomes of which 6 are associated in polysomes resulting in a frequency of 0.86. Random association probability was calculated by fractional abundance of each state in the dataset. For the plots showing the fold increase, the modelled mean and confidence interval lower and upper bounds were divided by the random association probability and displayed with logarithmic *y*-axis. Statistical significance for the fitted logistic parameters was determined with a two-sided Wald-test (as reported by mclogit) and used to annotate plots. *P* values were adjusted for multiple comparisons with the Hochberg method as implemented in R with p.adjust (method='hochberg').

Previously published data

We made use of previously published atomic models from the PDB (accession codes 5AJO, 4CXG, 4UJE, 6Y0G, 6Y57, 6GZ5, 6Z6L, 6Z6M, 5LZS, 4C0S, 5LZT, 5IZK, 6O85, 5LZZ, 6GZ3, 6GZ4, 6GZ5, 6SXO, 1BN5, 6W6L, 6ENY, 6S7O, 3JC2) and the AlphaFold Protein Structure Database

(AF-O00178, AF-P30101). Moreover, we used the following EM densities from the EMDB for analyses: EMDB-2904, EMDB-2908.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data generated in this study have been deposited at the Electron Microscopy Data Bank (www.ebi.ac.uk/emdb) under accessions EMD-15870, EMD-15871, EMD-15872, EMD-15873, EMD-15874, EMD-15875, EMD-15876, EMD-15877, EMD-15878, EMD-15879, EMD-15880, EMD-15884, EMD-15885, EMD-15886, EMD-15887, EMD-15888, EMD-15889, EMD-15890, EMD-15891, EMD-15892, EMD-15893 and the Protein Data Bank (www.rcsb.org) under accessions 8B6Z and 8B6L. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[68] partner repository with the dataset identifier PXD035475.

Code availability

Python code for polysome analysis is available at https://github.com/McHaillet/polysome-stats.

References

1. Voorhees, R. M. & Hegde, R. S. Toward a structural understanding of co-translational protein translocation. *Curr. Opin. Cell Biol.* **41**, 91–99 (2016).

2. Gemmer, M. & Förster, F. A clearer picture of the ER translocon complex. *J. Cell Sci.* **133**, jcs231340 (2020).

3. Voorhees, R. M. & Ramakrishnan, V. Structural basis of the translational elongation cycle. *Annu. Rev. Biochem.* **82**, 203–236

(2013).

4. Ramirez, A. S., Kowal, J. & Locher, K. P. Cryo-electron microscopy structures of human oligosaccharyltransferase complexes OST-A and OST-B. *Science* **366**, 1372–1375 (2019).

5. Van den Berg, B. et al. X-ray structure of a protein-conducting channel. *Nature* **427**, 36–44 (2004).

6. Voorhees, R. M. & Hegde, R. S. Structure of the Sec61 channel opened by a signal sequence. *Science* **351**, 88–91 (2016).

7. Braunger, K. et al. Structural basis for coupling protein transport and N-glycosylation at the mammalian endoplasmic reticulum. *Science* **360**, 215–219 (2018).

8. Zimmer, J., Nam, Y. & Rapoport, T. A. Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature* **455**, 936–943 (2008).

9. Egea, P. F. & Stroud, R. M. Lateral opening of a translocon upon entry of protein suggests the mechanism of insertion into membranes. *Proc. Natl Acad. Sci. USA* **107**, 17182–17187 (2010).

10. Fons, R. D., Bogert, B. A. & Hegde, R. S. Substrate-specific function of the translocon-associated protein complex during translocation across the ER membrane. *J. Cell Biol.* **160**, 529–539 (2003).

11. Nguyen, D. et al. Proteomics reveals signal peptide features determining the client specificity in human TRAP-dependent ER protein import. *Nat. Commun.* **9**, 3765 (2018).

12. Pfeffer, S. et al. Structure of the native Sec61 protein-conducting channel. *Nat. Commun.* **6**, 8403 (2015).

13. Pfeffer, S. et al. Dissecting the molecular organization of the translocon-associated protein complex. *Nat. Commun.* **8**, 14516 (2017).

14. Pfeffer, S. et al. Structure of the mammalian oligosaccharyl-transferase complex in the native ER protein translocon. *Nat. Commun.* **5**, 3072 (2014).

15. McGilvray, P. T. et al. An ER translocon for multi-pass membrane protein biogenesis. *eLife* **9**, e56889 (2020).

16. Hegde, R. S. & Keenan, R. J. The mechanisms of integral membrane protein biogenesis. *Nat. Rev. Mol. Cell Biol.* **23**, 107–124 (2021).

17. Tegunov, D., Xue, L., Dienemann, C., Cramer, P. & Mahamid, J. Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 Å in cells. *Nat. Methods* **18**, 186–193 (2021).

18. Pfeffer, S. et al. Structure and 3D arrangement of endoplasmic reticulum membrane-associated ribosomes. *Structure* **20**, 1508–1518 (2012).

19. Brandt, F., Carlson, L. A., Hartl, F. U., Baumeister, W. & Grunewald, K. The three-dimensional organization of polyribosomes in intact human cells. *Mol. Cell* **39**, 560–569 (2010).

20. Wells, J. N. et al. Structure and function of yeast Lso2 and human CCDC124 bound to hibernating ribosomes. *PLoS Biol.* **18**, e3000780 (2020).

21. Behrmann, E. et al. Structural snapshots of actively translating human ribosomes. *Cell* **161**, 845–857 (2015).

22. Loveland, A. B., Demo, G. & Korostelev, A. A. Cryo-EM of elongating ribosome with EF-Tu·GTP elucidates tRNA proofreading. *Nature* **584**, 640–645 (2020).

23. Shao, S. et al. Decoding mammalian ribosome–mRNA states by translational GTPase complexes. *Cell* **167**, 1229–1240.e1215 (2016).

24. Flis, J. et al. tRNA translocation by the eukaryotic 80S ribosome and the impact of GTP hydrolysis. *Cell Rep.* **25**, 2676–2688.e2677 (2018).

25. Budkevich, T. V. et al. Regulation of the mammalian elongation cycle by subunit rolling: a eukaryotic-specific ribosome rearrangement. *Cell* **158**, 121–131 (2014).

26. Crepin, T. et al. Mammalian translation elongation factor eEF1A2: X-ray structure and new features of GDP/GTP exchange mechanism in higher eukaryotes. *Nucleic Acids Res.* **42**, 12939–12948 (2014).

27. Song, H., Parsons, M. R., Rowsell, S., Leonard, G. & Phillips, S. E. Crystal structure of intact elongation factor EF-Tu from *Escherichia coli* in GDP conformation at 2.05 Å resolution. *J. Mol. Biol.* **285**, 1245–1256 (1999).

28. Xue, L. et al. Visualizing translation dynamics at atomic detail inside a bacterial cell. *Nature* **610**, 205–211 (2022).

29. Morse, J. C. et al. Elongation factor-Tu can repetitively engage aminoacyl-tRNA within the ribosome during the proofreading stage of tRNA selection. *Proc. Natl Acad. Sci. USA* **117**, 3610–3620 (2020).

30. Liu, W. et al. Labeled EF-Tus for rapid kinetic studies of pretranslocation complex formation. *ACS Chem. Biol.* **9**, 2421–2431 (2014).

31. Ieong, K. W., Uzun, U., Selmer, M. & Ehrenberg, M. Two proofreading steps amplify the accuracy of genetic code translation. *Proc. Natl Acad. Sci. USA* **113**, 13744–13749 (2016).

32. Wasserman, M. R., Alejo, J. L., Altman, R. B. & Blanchard, S. C. Multiperspective smFRET reveals rate-determining late intermediates of ribosomal translocation. *Nat. Struct. Mol. Biol.* **23**, 333–341 (2016).

33. Pavitt, G. D. & Ron, D. New insights into translational regulation in the endoplasmic reticulum unfolded protein response. *Cold Spring Harb. Perspect. Biol.* **4**, a012278 (2012).

34. Wild, K. et al. MetAP-like Ebp1 occupies the human ribosomal tunnel exit and recruits flexible rRNA expansion segments. *Nat. Commun.* **11**, 776 (2020).

35. Chitwood, P. J. & Hegde, R. S. An intramembrane chaperone complex facilitates membrane protein biogenesis. *Nature* **584**, 630–634 (2020).

36. Sundaram, A. et al. Substrate-driven assembly of a translocon for multipass membrane proteins. *Nature* **611**, 167–172 (2022).

37. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

38. Li, L. et al. Crystal structure of a substrate-engaged SecY protein-translocation channel. *Nature* **531**, 395–399 (2016).

39. Itskanov, S., Kuo, K. M., Gumbart, J. C. & Park, E. Stepwise gating of the Sec61 protein-conducting channel by Sec63 and Sec62. *Nat. Struct. Mol. Biol.* **28**, 162–172 (2021).

40. Wollenberg, M. S. & Simon, S. M. Signal sequence cleavage of peptidyl-tRNA prior to release from the ribosome and translocon. *J. Biol. Chem.* **279**, 24919–24922 (2004).

41. Smalinskaitė, L., Kim, M. K., Lewis, A. J. O., Keenan, R. J. & Hegde, R. S. Mechanism of an intramembrane chaperone for multipass membrane proteins. *Nature* **611**, 161–166 (2022).

42. Kriegler, T., Lang, S., Notari, L. & Hessa, T. Prion protein translocation mechanism revealed by pulling force studies. *J. Mol. Biol.* **432**, 4447–4465 (2020).

43. Wiedmann, M., Goerlich, D., Hartmann, E., Kurzchalia, T. V. & Rapoport, T. A. Photocrosslinking demonstrates proximity of a 34 kDa membrane protein to different portions of preprolactin during translocation through the endoplasmic reticulum. *FEBS lett.* **257**, 263–268 (1989).

44. Karki, S. et al. Molecular view of ER membrane remodeling by the Sec61/TRAP translocon. Preprint at *bioRxiv* https://doi.org/10.1101/2022.09.30.510141 (2022).

45. Aebi, M. *N*-linked protein glycosylation in the ER. *Biochim. Biophys. Acta* **1833**, 2430–2437 (2013).

46. Gidalevitz, T., Stevens, F. & Argon, Y. Orchestration of secretory protein folding by ER chaperones. *Biochim. Biophys. Acta* **1833**, 2410–2424 (2013).

47. Tian, G., Xiang, S., Noiva, R., Lennarz, W. J. & Schindelin, H. The crystal structure of yeast protein disulfide isomerase suggests cooperativity between its active sites. *Cell* **124**, 61–73 (2006).

48. Frickel, E. M. et al. TROSY-NMR reveals interaction between ERp57 and the tip of the calreticulin P-domain. *Proc. Natl Acad. Sci. USA* **99**, 1954–1959 (2002).

49. Oliver, J. D., Roderick, H. L., Llewellyn, D. H. & High, S. ERp57 functions as a subunit of specific complexes formed with the ER lectins calreticulin and calnexin. *Mol. Biol. Cell* **10**, 2573–2582 (1999).

50. Mohorko, E. et al. Structural basis of substrate specificity of human oligosaccharyl transferase subunit N33/Tusc3 and its role in regulating protein *N*-glycosylation. *Structure* **22**, 590–601 (2014).

51. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol* **26**, 1367–1372 (2008).

52. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).

53. Hagen, W. J., Wan, W. & Briggs, J. A. Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *J. Struct. Biol.* **197**, 191–198 (2016).

54. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods* **16**, 1146–1152 (2019).

55. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).

56. Hrabe, T. et al. PyTom: a python-based toolbox for localization of macromolecules in cryo-electron tomograms and subtomogram analysis. *J. Struct. Biol.* **178**, 177–188 (2012).

57. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).

58. Varadi, M. et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).

59. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D* **75**, 861–877 (2019).

60. Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. D* **74**, 519–530 (2018).

61. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).

62. Evans, R. et al. Protein complex prediction with AlphaFold-Multimer. Preprint at *bioRxiv* https://doi.org/10.1101/2021.10.04.463034 (2021).

63. Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. *J. Struct. Biol.* **157**, 281–287 (2007).

64. Lopez-Blanco, J. R. & Chacon, P. iMODFIT: efficient and robust flexible fitting based on vibrational analysis in internal coordinates. *J. Struct. Biol.* **184**, 261–270 (2013).

65. Williams, C. J. et al. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).

66. Ashkenazy, H. et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016).

67. Elff, M. Social divisions, party positions, and electoral behaviour. *Elect. Stud.* **28**, 297–308 (2009).

68. Perez-Riverol, Y. et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res* **50**, D543–D552 (2022).

69. Chen, Y. et al. Detection and identification of macromolecular complexes in cryo-electron tomograms using support vector machines. In *2012 9th IEEE International Symposium on Biomedical Imaging* 1373–1376 (IEEE, 2012).

70. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).

Acknowledgements

This work was supported by the European Research Council under the European Union's Horizon2020 Program (ERC Consolidator Grant Agreement 724425–BENDER) and the Nederlandse Organisatie voor Wetenschappelijke Onderzoek (Vici 724.016.001 to F.F., Veni 212.152 to J.F., research program TA 741.018.201 to R.A.S. and F.F., and National Roadmap for Large-Scale Research Infrastructure NEMI 184.034.014). We thank G. van der Schot, M. Vanevic and R. Englmeier for help with data processing; R. Hermsen for advice on statistical analysis; and S. Pfeffer, B. Janssen, M. Feige, E. Schmitt, Y. Mechulam, S. Lang and R. Zimmermann for stimulating discussions and critical comments on the manuscript.

Author information

Authors and Affiliations

1. Structural Biochemistry, Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands

   Max Gemmer, Marten L. Chaillet, Joyce van Loenhout, Rodrigo Cuevas Arenas, Dimitrios Vismpas, Mariska Gröllers-Mulderij, Stuart C. Howes, Juliette Fedry & Friedrich Förster

2. Thermo Fisher Scientific, Eindhoven, The Netherlands

   Fujiet A. Koh & Abhay Kotecha

3. Biomolecular Mass Spectrometry and Proteomics Group, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands

   Pascal Albanese & Richard A. Scheltema

4. Netherlands Proteomics Center, Utrecht University, Utrecht University, Utrecht, The Netherlands

   Pascal Albanese & Richard A. Scheltema

Contributions

M.G., J.F. and F.F. conceived the project. M.G. performed microsome sample preparation, cryo-ET data acquisition and image analysis. M.G. and J.F. carried out model building. M.L.C. analysed polysomes. J.v.L., S.C.H. and J.F. performed in situ cryo-ET of human cells. R.C.A. and M.G.-M. cloned CRISPR–Cas9 constructs and generated monoclonal cells. D.V., M.G., P.A. and R.A.S. analysed samples using mass spectrometry. J.F. performed cryo-EM SPA sample preparation and data processing, F.A.K. and A.K. collected cryo-EM SPA data. M.G., M.L.C., J.F. and F.F. analysed the data and wrote the manuscript.

Corresponding authors

Correspondence to Juliette Fedry or Friedrich Förster.

Ethics declarations

Competing interests

A.K. and F.A.K. are employees of Thermo Fisher Scientific. The other authors declare no competing interests.

Peer review

Peer review information

Nature thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Cryo-ET data analysis workflow.

Template matching in PyTom[69] generates candidates for ribosomal particles, which are further analyzed in RELION[57] and M[17]. Initial coarse 3D classification allowed removal of false positives, poorly aligned particles, and isolated LSUs. (**A**) The remaining ~135,000 80S ribosome subtomograms were subjected to focused classification on the area at the ribosomal tunnel exit (mask 1). Repeated classification is required to distinguish subtle differences of Sec61-multipass-, Sec61-multipass-TRAP translocon, and Sec61-TRAP. (**B**) The center of the reconstruction of the ribosome-Sec61-TRAP-OSTA population was shifted to the center of the translocon. After refinement, recentered subtomograms were subjected to 3D classification focused on a luminal mask near OSTA (mask 4). (**C**) To obtain the best statistics for analysis of ribosomal processing states all

subtomograms were pooled again. The particles were hierarchically classified, first according to the rotation state of the SSU (mask 2) and then further focused using masks including the tRNA and eEF binding sites (mask 3). A minor population of <2k particles could not be assigned unambiguously to a translation state (ND = not defined). (**D**) Previously annotated particles from classification focused on the translocon (**A**) were extracted from classes obtained by classification of ribosomal intermediate states (**C**).

Extended Data Fig. 2 Identification of ribosomal intermediate states.

Large ribosomal subunits of models or maps of previously characterized intermediate states were fitted into our reconstructions from Fig. 1A, of which we only show the tRNAs and elongation factors for clarity. Structures of mRNAs, tRNAs, elongation factors and the small ribosomal subunit from the models indicated by their PDB or EMDB codes are superposed onto the respective segmented densities from our reconstructions.

Extended Data Fig. 3 Neighborhood analysis of ER membrane-bound and soluble ribosomes and their intermediate states.

(**A**) Side view (top panels) and top view (bottom panels) of filtered reconstructions of ER-membrane bound, soluble and hibernating ribosome populations depicted at low contour level. Densities of leading and trailing ribosome neighbors are visible adjacent to the centered ribosome. (**B**) Neighborhood analysis illustrates the arrangement of ribosomes and is consistent with the subtomogram averages from (**A**). Neighborhood analysis was performed in 3D, whereas 2D heat maps show the results projected onto a plane parallel to the membrane. (**C**) Masks were generated in 3D from results of the neighborhood analysis of membrane-bound and soluble populations combined. (**D**) Columns represent the modelled mean neighbor probability with 95% confidence interval as error bars analysis based on the neighborhood analysis from (**B,C**) for each ribosomal intermediate state. Statistics determined from n = 132,371 ribosomes with the 869 tomograms included as a random effect. The random association probability (gray hatched bars) is the overall abundance of the ribosome populations. (**E**)

Columns represent the mean logarithmic fold increase of observed vs. random probability with 95% confidence interval as error bars of the data from (**D**).

Extended Data Fig. 4 Ribosome states *in situ* and comparison to *ex vivo* abundances.

(**A**) Central slice (thickness 1.7 nm) of representative tomograms of cryo-FIB milled HEK293, U2OS and HeLa cells. Scale bar: 100 nm. (**B**) Segmented representation of tomograms from (**A**). Subtomogram averages of the ribosome were mapped back into the reconstruction and color-coded according to their ribosomal state. (**C**) Ribosomal states obtained by 3D classification of *in situ* data. (**D**) Neighborhood analysis of the intermediate states from (**C**). (**E**) Distribution of ribosomal states from soluble or membrane-bound ribosomes. Statistics determined from n = 132,371 ribosomes with 869 tomograms modeled as random effect. Stacked columns show the modelled mean with the 95% confidence interval as error bars. (**F**) Distribution of ribosomal states from 3 separate ER vesicles preparations (*ex vivo* - ER #1-3), *in situ* data, and cytosolic polysomes from Behrmann et al[21]. n(ER #1) = 132,731 particles in 869 tomograms, n(ER #2) = 6,101 particles in 31 tomograms, n(ER #3) = 3,836 particles in 58 tomograms, each from 1 experiment, n(in situ) = 5,351 (HEK293 = 2,965, U2OS = 374, HeLa = 2,012) particles in 27 tomograms from 3 independent experiments.

Extended Data Fig. 5 Identification of elongation factor-bound ribosomal intermediate states.

(**A**) Superposition of the decoding-sampling (4CXG+4UJE), decoding-recognition (5LZS) and post-decoding (EMDB-2908) state (dark grey cartoon representations) onto our reconstruction (semi-transparent colored maps) of the ribosome-bound eEF1a-tRNA ternary complex. Arrows indicate structural differences. (**B**) Close-up of the decoding center of the decoding-recognition state (5LZS) superposed onto our segmented reconstructions (semi-transparent maps) of our decoding state (left) or the subsequent classical pre state (right) for comparison. Densities of the nucleobases A1824 and A1825 are clearly visible in the flipped-out conformation in the classical pre state (right) but flipped-in in the decoding

state (left), indicating that tRNA recognition has not yet occurred. tRNA, mRNA, and 18S rRNA segment h44 were segmented and tRNAs were clipped for better overview. (**C**) Comparison of eEF1a and structurally related candidates fitted into the segmented density of the classical pre+ state. Arrowheads indicate structural differences. (**D**) Structure of eEF1A in extended conformation (4C0S) fitted into the segmented density of the classical pre+ state. Domain 1, 2 and 3 (D1-3) were fitted individually.

Extended Data Fig. 6 Single particle analysis of the ribosome in the classical pre+ state.

(**A**) Comparison of cryo-ET and SPA reconstructions of the ribosome in the classical pre+ state filtered to local resolution. Ribosomes were clipped in top views (bottom panels). (**B**) SPA reconstruction color-coded according to local resolution. (**C**) Close-up view of eEF1a color-coded according to local resolution explained in the color bar. (**D**) Refined atomic model of eEF1a placed into the SPA density map. Domains 1-3 (D1-3) are indicated. (**E**) Segments of eEF1a superposed on density maps with well-resolved side chains. (**F**) Refined model of eEF1a fitted into the locally refined reconstruction of domain 3. The SRL is not depicted for clarity. (**G**) Candidate GTPases fitted into the high-resolution density. The SRL binding site of domain 3 is displayed. (**H**) Interaction site of eEF1a with the SRL of the LSU. (**H**) Same view as in (**H**) with the density map.

Extended Data Fig. 7 Neighbor probability analysis of soluble and ER translocon populations.

(**A**) Central slices from representative filtered tomograms of ER-derived vesicles. ER (endoplasmic reticulum), V (vesicle), C (carbon support). (**B**) Segmented representation of tomograms from (**A**), including the ER membrane (grey), carbon support (black) and subtomogram averages of different ribosome populations mapped back into the tomogram. Ribosomes are color-coded according to their binding partners at the exit tunnel: soluble (blue), OSTA-translocon (red), TRAP-translocon (green), multipass-translocon (yellow), unassigned (grey); large ribosomal subunit (LSU, lighter shade), small ribosomal subunit (SSU, darker shade). (**C**) Probability of encountering soluble or ER-associated ribosomes from as leading or

trailing neighbor. The black circles show the modelled mean with the 95% confidence interval as error bars fitted to n = 134,350 ribosomes with the 869 tomograms included as a random effect. The small scattered points represents the frequencies of events per tomogram. The random association probability (bright red lines) is the overall abundance of the ribosome populations corrected for unoccupied positions. Neighbors are defined as ‘unoccupied’ if there is no particle in the defined neighborhood mask or its potential neighbor (e.g., a particle must have a trailing neighbor, which has this particle as a leading neighbor). (**D**) Columns represent the mean fold increase of observed vs random probability with 95% confidence interval as error bars of the data from (**C**).

Extended Data Fig. 8 Reconstruction of the Sec61-TRAP-OSTA-translocon.

(**A**) Ribosome- and translocon-centered reconstruction of the ribosome-Sec61-TRAP-OSTA-translocon color-coded by local resolution (color bar in Å). Centers of the respective reconstructions are indicated. (**B**) FSC curves of the ribosome- and translocon-centered reconstructions of the ribosome-Sec61-TRAP-OSTA-translocon. (**C**) Examples of 60S ribosomal proteins and 28S rRNA fitted into the ribosome-centered reconstruction filtered to local resolution of up to 3.5-Å. (**D**) Cryo-EM structures of Sec61 (3JC2) fitted into the translocon-centered reconstruction. (**E**) Density of the nascent chain (NC, light-yellow) is visible at the ribosomal tunnel exit, the Sec61 pore and in the lateral gate as signal peptide (SP, light-yellow). The front side of the ribosome and membrane were clipped for visualization purposes. (**F**) Close-up of the Sec61 plug placed into the density of the translocon-centered reconstruction. (**G**) Superposition of the plug in the closed (cyan, 3J7Q) and open (blue) conformation.

Extended Data Fig. 9 Model building of the TRAP complex.

(**A**) Prediction model of TRAP (P43307, P43308, Q9UNL2, P51571) obtained by Colabfold (v1.4)[70] using MMseqs2 and Alphafold2-multimer (v2)[62] color-coded according to predicted local distance difference test (pLDDT) score. Signal peptides were removed prior to prediction. (**B**) Sequence coverage obtained by sequence alignments generated by

MMseqs2. (**C**) pLDDT scores per position of five model predictions. (**D**) Predicted aligned error (PAE) of five model predictions. (**E**) Prediction models of TRAPαβδ placed into the density of the locally filtered translocon-centered reconstruction. (**F**) Alphafold models of TRAPβγδ placed into the segmented density of the locally filtered ribosome-centered reconstruction. (**G**) Additional densities which are not explained by the prediction models reside near disordered terminal regions (white arrowhead) or glycosylation sites of TRAPαβ indicating partially ordered glycans (black arrowheads). Asparagine residues are displayed as ball/stick models and annotated according to residue number. (**H**) Sequence conservation score plotted onto the surface of TRAP subunits (blue: high conservation, orange: low conservation). Evolutionary conserved residues reside primarily at the interface areas, whereas peripheral residues are variable. The luminal TRAPα, TRAPβ, and TRAPδ domains possess large interaction interfaces (TRAPα-TRAPβ: 695 $Å^2$, TRAPβ-TRAPδ: 985 $Å^2$). (**I**) Top, back and side view of the reconstruction of the Sec61-TRAP-OSTA-translocon (top panels). Semi-transparent densities originate from residual membrane signal. Models generated from the density map at the same view (bottom panels).

Extended Data Fig. 10 Native OSTA and its accessory factors.

(**A**) View from cytosol (top) and side view (bottom) of the OSTA complex (PDB 6S7O, AlphaFold P04844) fitted into the segmented map of the translocon-centered reconstruction of the OSTA-translocon. (**B**) AlphaFold model of RPN2 (P04844). The model is color-coded according to confidence score as indicated. (**C**) Close-up view of the N-terminal domain (NTD) of the RPN2 prediction model fitted into the reconstruction as in (A). (**D**) Side view of the OSTA-translocon opposite to the lateral gate. (**E,F**) Close-up side view (**E**) and top view from the cytosol (**F**) of T1 intercalated between TMHs of STT3a and TRAPα. (**G,H**) Membrane-resident translocon components (same view as in (**F**)) of the ribosome-centered reconstructions of the Sec61-TRAP-OSTA-translocon (**G**) and the Sec61-TRAP-translocon (**H**) filtered to a resolution of 15 Å. (**I–K**) Reconstructions of the OSTA-translocon without (**I**) or with accessory factor L1 (**J**) or L2 (**K**) color-coded according to local resolution as indicated. (**L**) FSC curves of the reconstructions from (**I–K**). (**M**) Models of L2-candidate proteins PDIA3 (6ENY) and PDIA5 (Q14554). Catalytic (a, a’) and non-catalytic (b, b’)

thioredoxin domains are indicated. (**N,O**) PDI domains a and b fitted into the reconstruction of OSTA-L2. (**P**) Sequence conservation plotted onto the surface model of the RPN2 NTD. Highly conserved residues reside at the binding site of the a'-domain of PDI or other OST subunits. (**Q**) Close-up view of the interaction site of STT3A and L2-1.

Supplementary information

Supplementary Information

This file contains Supplementary Figs. 1–7 and Supplementary Tables 1 and 2.

Reporting Summary

Supplementary Video 1

The elongation cycle. Experimentally observed elongation intermediates are arranged in the model of the human elongation cycle.

Supplementary Video 2

Conformational changes of eEF1a on ribosomes. Morph illustrating the conformational change of eEF1a from the compact conformation (decoding-recognition state, 5LZS) to the extended conformation (classical pre+ state).

Peer Review File

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's

Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Gemmer, M., Chaillet, M.L., van Loenhout, J. *et al.* Visualization of translation and protein biogenesis at the ER membrane. *Nature* **614**, 160–167 (2023). https://doi.org/10.1038/s41586-022-05638-5

- Received: 27 January 2022
- Accepted: 07 December 2022
- Published: 25 January 2023
- Issue Date: 02 February 2023
- DOI: https://doi.org/10.1038/s41586-022-05638-5

Share this article

Anyone you share the following link with will be able to read this content:

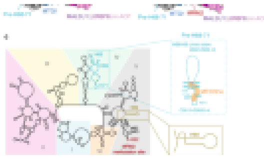
Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

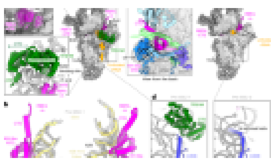
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05638-5



Structural basis for late maturation steps of the human mitoribosomal large subunit

16 June 2021

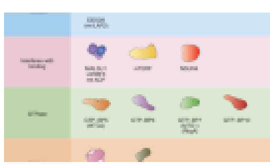
Miriam Cipullo, Genís Valentín Gesé, … Joanna Rorbach



Mechanism of mitoribosomal small subunit biogenesis and preinitiation

08 June 2022

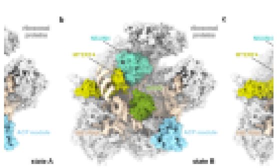
Yuzuru Itoh, Anas Khawaja, … Alexey Amunts



Mitoribosome assembly comes into view

11 August 2021

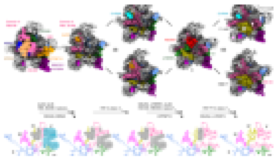
Marie Sissler & Yaser Hashem



[Stepwise maturation of the peptidyl transferase region of human mitoribosomes]

16 June 2021

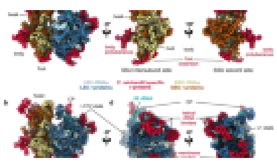
Tea Lenarčič, Mateusz Jaskolowski, … Nenad Ban



[A distinct assembly pathway of the human 39S late pre-mitoribosome]

27 July 2021

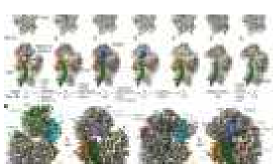
Jingdong Cheng, Otto Berninghausen & Roland Beckmann



[How to build a ribosome from RNA fragments in Chlamydomonas mitochondria]

09 December 2021

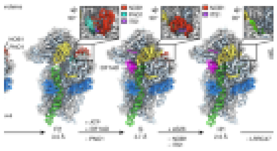
Florent Waltz, Thalia Salinas-Giegé, … Yaser Hashem



[Visualizing late states of human 40S ribosomal subunit maturation]

06 June 2018

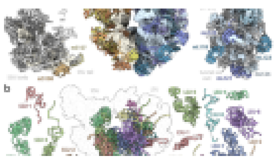
Michael Ameismeier, Jingdong Cheng, … Roland Beckmann



Structural basis for the final steps of human 40S ribosome maturation

18 November 2020

Michael Ameismeier, Ivo Zemp, … Roland Beckmann



Structure of a mitochondrial ribosome with fragmented rRNA in complex with membrane-targeting elements

17 October 2022

Victor Tobiasson, Ieva Berzina & Alexey Amunts

- Article
- Open Access
- Published: 08 December 2022

Principles of mitoribosomal small subunit assembly in eukaryotes

- Nathan J. Harper,
- Chloe Burnside &

- Sebastian Klinge

Nature **volume 614**, pages 175–181 (2023)

- 5597 Accesses
- 104 Altmetric
- Metrics details

Abstract

Mitochondrial ribosomes (mitoribosomes) synthesize proteins encoded within the mitochondrial genome that are assembled into oxidative phosphorylation complexes. Thus, mitoribosome biogenesis is essential for ATP production and cellular metabolism[1]. Here we used cryo-electron microscopy to determine nine structures of native yeast and human mitoribosomal small subunit assembly intermediates, illuminating the mechanistic basis for how GTPases are used to control early steps of decoding centre formation, how initial rRNA folding and processing events are mediated, and how mitoribosomal proteins have active roles during assembly. Furthermore, this series of intermediates from two species with divergent mitoribosomal architecture uncovers both conserved principles and species-specific adaptations that govern the maturation of mitoribosomal small subunits in eukaryotes. By revealing the dynamic interplay between assembly factors, mitoribosomal proteins and rRNA that are required to generate functional subunits, our structural analysis provides a vignette for how molecular complexity and diversity can evolve in large ribonucleoprotein assemblies.

Main

The mitoribosome, which is present in the mitochondria of eukaryotic cells, translates mitochondrial mRNAs that predominantly encode components of oxidative phosphorylation complexes. This molecular machine comprises rRNAs (12S and 16S in humans, and 15S and 21S in yeast) encoded in the mitochondrial genome that associate with predominantly nuclear-encoded

mitochondrial ribosomal proteins (mito r-proteins) in the mitochondrial matrix to form the mitoribosome (55S in human and 74S in yeast). Biogenesis of the mitoribosome requires *trans*-acting assembly factors that chaperone rRNA folding and guide mito r-protein incorporation[2,3,4,5,6]. The central role of mitoribosomes to cellular metabolism is highlighted by several human diseases caused by mutations in either mito r-proteins or assembly factors[6,7,8,9,10]. A structural understanding of the principles of mitoribosome assembly is therefore crucial to define the molecular underpinnings of these processes and their associated human diseases.

Although mitochondrial large subunit assembly has been extensively studied in kinetoplastids[7,8] and human cells[9,10,11,12,13,14], our structural understanding of mitochondrial small subunit (mtSSU) assembly is so far limited to *Trypanosoma brucei*[15,16], an extreme example in terms of evolutionary divergence, and late stages of mammalian mitoribosome assembly[17]. We currently lack fundamental insights into the initial steps of mtSSU assembly in key organisms such as *Saccharomyces cerevisiae* and *Homo sapiens*.

Despite substantial structural divergence across ribosomes from various species[2,3,4,5,6,18,19,20,21], all SSU assembly systems must overcome several common rRNA folding hurdles to produce functional subunits. These include the formation of the decoding centre, the key functional module of the SSU, which must be formed in a regulated manner to generate subunits exhibiting high translational fidelity. In addition, the 5′ and 3′ ends of the rRNA need to be generated and integrated into the maturing particle. Simultaneously, premature engagement of SSU assembly intermediates with mature large ribosomal subunits must be prevented to ensure a functional ribosome pool. Although different solutions have evolved for these hurdles in bacterial and eukaryotic systems[22,23], little is known about how yeast and human cells achieve control over these initial assembly steps and how these systems have evolved to catalyse the formation of functional mitoribosomal SSUs that are structurally diverse. To address this, we solved high-resolution structures of six native assembly intermediates from human cells and three native assembly intermediates from yeast cells by cryo-electron microscopy (cryo-EM), revealing several fundamental principles of mtSSU assembly. First, stepwise activities of molecular switches control early rRNA folding

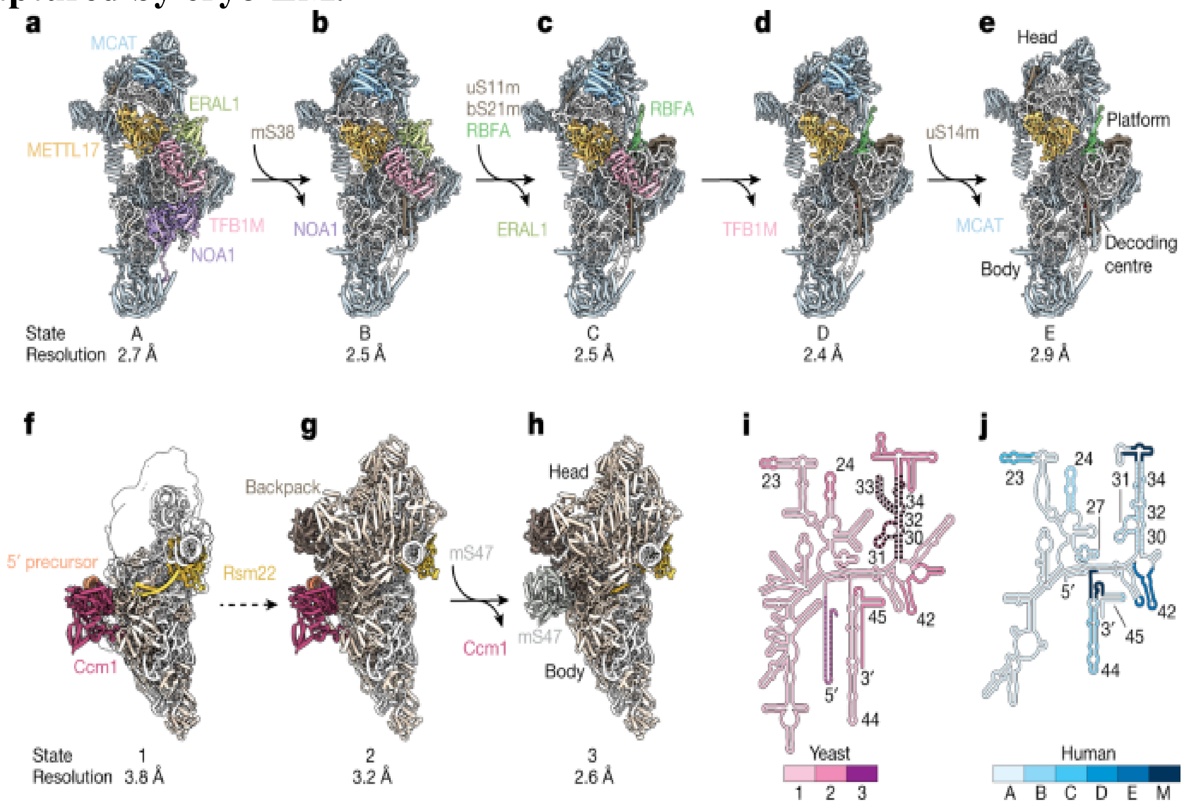
events that lay a foundation for the formation of a functional decoding centre. Second, integration of the rRNA 3′ end facilitates compaction of the functional rRNA core, with yeast and human systems each having evolved distinct solutions to do so. Third, recognition and processing of 5′ pre-rRNA is a unique feature of the yeast mtSSU assembly pathway, and is carried out by distinct assembly factors. Fourth, a conserved assembly factor orchestrates maturation of the head domain and prevents premature engagement of mRNA and the mitochondrial large subunit. Together, these principles shed light on the evolution of mtSSU assembly pathways and how species-specific adaptations are used to generate both molecular complexity and diversity.

Cryo-EM snapshots of mtSSU maturation

To capture human mtSSU assembly intermediates, we used an established CRISPR-based biallelic tagging method to generate a stable HEK293F cell line endogenously expressing the poorly understood assembly factor METTL17 C-terminally tagged with green fluorescent protein (GFP), and confirmed that tagging does not affect mitochondrial translation[24] (Supplementary Fig. 2). Native complexes were isolated by affinity purification, and liquid chromatography with tandem mass spectrometry (LC–MS/MS) revealed the presence of co-purified mtSSU factors and mito r-proteins (Supplementary Data 1). Using cryo-EM, we identified six discrete assembly states (A, B, C, D, E and C*) of the human mtSSU through global and focused classification strategies and solved their structures to resolutions ranging from 2.4 to 3 Å (Fig. 1a–e and Supplementary Figs. 3–19). These intermediates can be ordered into an assembly pathway that rationalizes the maturation status of rRNA elements as well as the presence of assembly factors and mito r-proteins. Although our assignment of states A–E suggests a linear assembly pathway of intermediates bound by METTL17, we cannot exclude the possibility of alternative pathways present in the cell, including those in which certain proteins undergo multiple cycles of binding and dissociation. In particular, the identification of the less populated assembly state C* suggests uncoupling of head and body assembly mechanisms during maturation, supporting the notion that mtSSU assembly does not occur in a strictly linear manner, in line with the parallel assembly pathways previously characterized

for bacterial SSU assembly[25] (Supplementary Figs. 9 and 18). These reconstructions reveal the structures and functions of six assembly factors in the context of assembly: the putative methyltransferase METTL17, the GTPases NOA1 (refs. [26,27]) and ERAL1 (refs. [28,29]), the methyltransferase TFB1M[30], the rRNA chaperone RBFA[31], and MCAT (malonyl-CoA-acyl carrier protein transacylase), a protein so far only implicated in mitochondrial fatty acid metabolism[32].

Fig. 1: Chronology of mitoribosomal SSU assembly intermediates captured by cryo-EM.



a–**e**, Atomic models of five human mtSSU assembly intermediates along the assembly pathway. Assembly factors are coloured and changes in protein composition between states are noted. The overall resolution of each cryo-EM reconstruction is noted below each state (see also Supplementary Figs. 2–18). **f**–**h**, Atomic models of three yeast mtSSU assembly intermediates along the assembly pathway. Assembly factors are coloured and changes in protein composition between states are noted. For state 1 (**f**), the model is shown with the corresponding cryo-EM map filtered to local resolution. The overall resolution of each cryo-EM reconstruction is noted below each state

(see also Supplementary Figs. 20–25). The dashed arrow between **f** and **g** indicates multiple maturation events. **i**,**j**, rRNA secondary structure diagrams of the yeast (**i**) and human (**j**) mtSSU. rRNA elements, which become ordered during the assembly pathway, are coloured according to the state in which the segment becomes ordered, with M corresponding to the mature structure (Protein Data Bank (PDB): 3J9M). Remodelled rRNA is indicated by hashes and processed rRNA is indicated by dots.

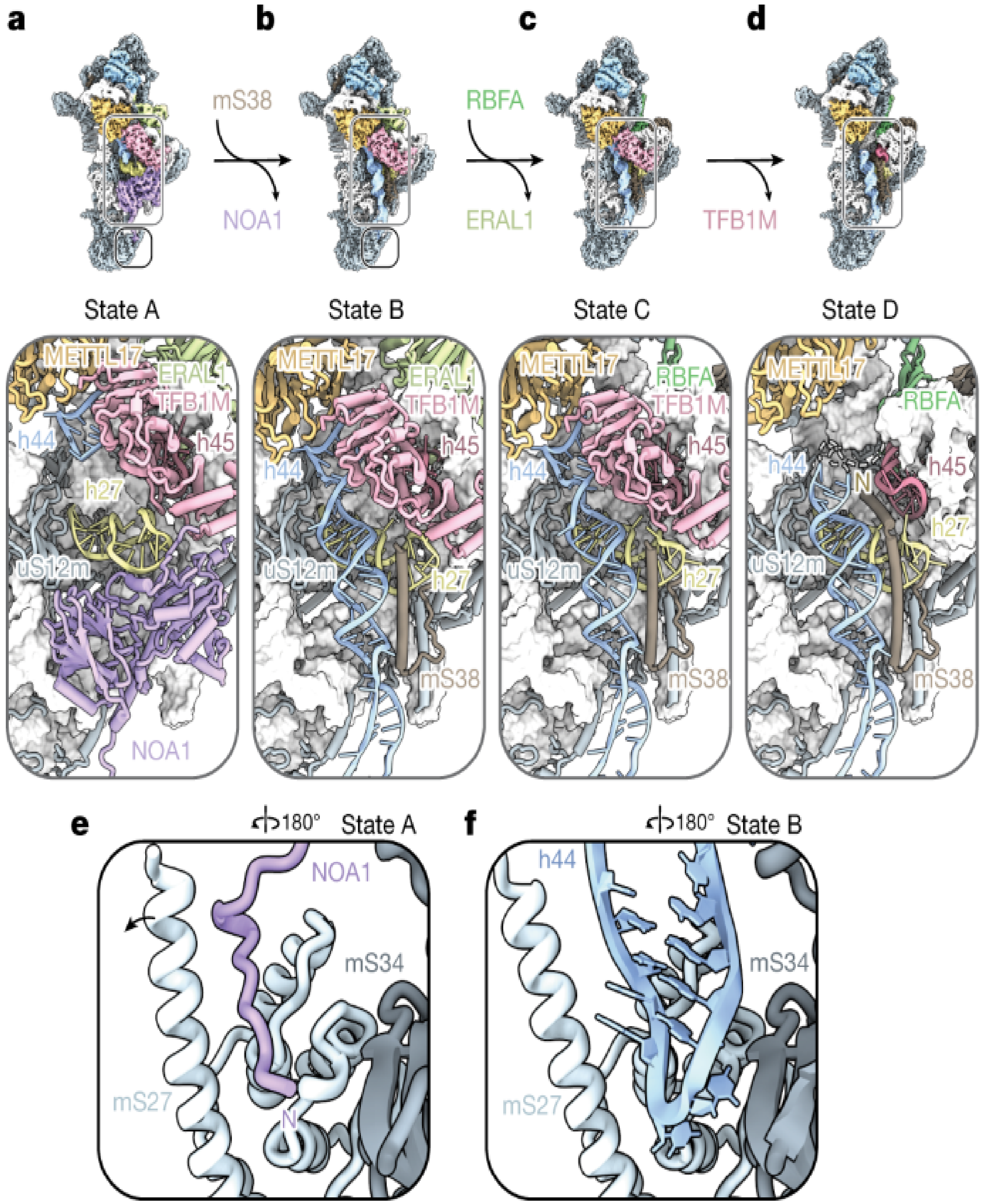
To gain insight into mtSSU biogenesis in yeast, we purified assembly intermediates using two baits (Supplementary Fig. 20). First, by using the yeast-specific factor Ccm1 (ref. [33]) as bait for affinity purification, we obtained two early assembly intermediates, states 1 and 2, which contain both Ccm1 and the METTL17 homologue Rsm22 (Fig. 1f,g and Supplementary Figs. 21,22 and 24–26). State 1 represents the earliest intermediate missing significant modules of mito r-proteins and exhibits high levels of conformational flexibility in the head region. Second, we purified particles by tagging Rsm22 and solved the structure of a Rsm22-bound intermediate, referred to as state 3 (Fig. 1h and Supplementary Figs. 23–26). LC–MS/MS analysis of both yeast complexes further revealed the presence of co-purified mtSSU assembly factors Mtg3 (homologous to human NOA1) and Rmd9, a protein so far only implicated in the protection of 3′ ends of mitochondrial mRNAs[34,35] (Supplementary Data 2 and 3 and Supplementary Fig. 20). The presence of all four assembly factors (Mtg3, Ccm1, Rmd9 and Rsm22) in an additional assembly intermediate was also confirmed by LC–MS/MS when Mtg3 was used as bait for affinity purification (Supplementary Data 4). However, these particles were not amenable to structural characterization, presumably owing to increased flexibility.

Using our structures of yeast and human mtSSU assembly intermediates, we can rationalize differences in rRNA and mito r-protein composition between yeast and human mtSSUs in the context of assembly, showing how differences are mirrored with contrasting local assembly mechanisms (Extended Data Fig. 1). However, despite the significant divergence in mitochondrial rRNA in these organisms, our reconstructions highlight a shared chronology in which key rRNA segments are remodelled or become ordered as a function of different assembly stages (Fig. 1i,j).

Guided formation of the decoding centre

The decoding centre is the key functional region of the SSU, which probes mRNA–tRNA interactions to ensure translational fidelity. The controlled formation of the decoding centre is therefore critical for the assembly of functional SSUs. The human assembly system uses two GTPases, NOA1 and ERAL1, which act as molecular switches that promote the first steps in the formation of the decoding centre and the platform (Fig. 2). In state A, NOA1 regulates docking of helix 44 (h44) through two mechanisms. First, an N-terminal extension of NOA1 blocks access to the h44-binding site at the tail of the mtSSU, which is occupied by h44 in state B (Fig. 2e,f). Second, the globular domains of NOA1 hinder h44 docking at the subunit interface while also sequestering h27 away from its mature position (Fig. 2a and Extended Data Fig. 2a,b). Dissociation of NOA1 allows for docking of h44 and its stabilization through binding of a helical segment of mS38 as well as repositioning of h27 to form the base of the platform (Fig. 2b).

Fig. 2: Initial steps of formation of the decoding centre in the human mtSSU.



a–**d**, Cryo-EM maps of human states A–D (top), with models highlighting stepwise maturation of h44 and the decoding centre (bottom). **e**, Zoomed-in view of the tail region in state A, showing how the N-terminal tail of NOA1 prevents h44 docking. **f**, Zoomed-in view of the tail region in state B, showing accommodation of h44 after NOA1 dissociation.

During eukaryotic and bacterial ribosome assembly, GTPases frequently use conformational changes coupled to nucleotide hydrolysis to promote directional assembly of rRNA and protein modules[36,37]. Within NOA1, we found a vacant GTP-binding site that is partially occluded by a helical insertion of NOA1, which prevents the switch I motif from occupying a conformation competent for GTP binding (Extended Data Fig. 2d). With binding of NOA1 to RNA[26] and stimulation of NOA1-mediated GTP hydrolysis by RNA having previously been shown[27], our structure of NOA1 in physiological context with a maturing mtSSU provides insight into how this subfamily of GTPases, which include the bacterial ribosome assembly factor YqeH, couple a GTPase cycle to rRNA conformational changes in mitochondrial and bacterial SSU assembly[37,38].

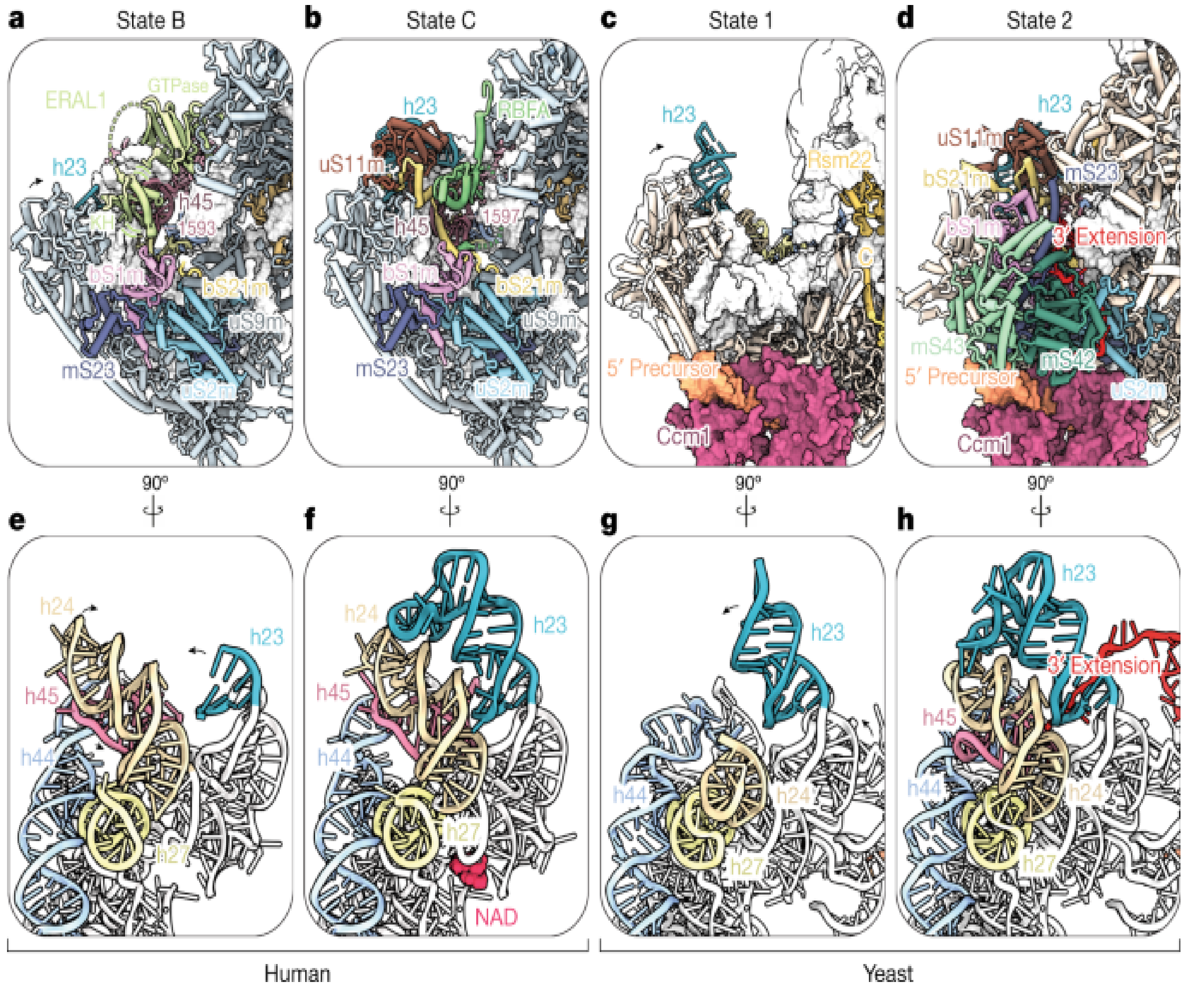
Positioned next to NOA1 is the methyltransferase TFB1M, which has a bacterial homologue (KsgA), that methylates adjacent adenosine residues in h45 of the 16S rRNA[39]. In addition to its role in rRNA methylation, discussed in detail in the next section, TFB1M fulfils a second role during mtSSU assembly as one-half of a molecular clamp with METTL17, which stabilizes a segment of h44 that interacts with mRNA during translation[40] (Extended Data Fig. 3a,b). Furthermore, TFB1M initially sequesters h45, preventing h45 from interacting with and stabilizing the mature conformation of h44 until TFB1M methylates two residues in h45 and departs during the transition from state C to state D (Fig. 2a–d and Extended Data Fig. 3c–e). As a result, this molecular clamp inhibits the premature folding of decoding centre rRNA elements and forms an assembly checkpoint in which h45 methylation is required for maturation of the decoding centre. Dissociation of NOA1 and TFB1M during assembly progressively generates the binding site for mS38, which in state D locks the relative orientation of h44, h27 and h45 in a near-mature conformation (Fig. 2a–d). Furthermore, in states C and D, the rRNA linkers connecting h44 to h45 and h28, which also have direct roles in mRNA decoding during translation, remain flexible due to the presence of the 3′ rRNA-binding assembly factor RBFA, as observed in bacteria[23]. Thus, the functional interplay between the GTPase NOA1, the methyltransferase TFB1M and the RNA chaperone RBFA guide and control the formation of the decoding centre such that h44 docking, stabilization and folding of rRNA linkers occur in a highly controlled and stepwise manner.

The final steps in the formation of the decoding centre require positioning of h31 proximal to the tRNA-binding P site. In human states E and C*, access of h31 to the decoding centre is blocked by assembly factor METTL17, a mechanism that is also present in the yeast system, with the METTL17 homologue Rsm22 performing the same function (Extended Data Fig. 4a–f). Although human mS38 binding occurs in a progressive manner and is present in states B–E, yeast mS38 is not yet bound in state 3, suggesting that mS38 binding and final rearrangements to the decoding site are initiated upon dissociation of Rsm22 (Extended Data Fig. 4g–j).

Platform and 3′ end rRNA stabilization

The correct formation of the platform and integration of the 3′ end of the mtSSU rRNA is critical for the stability of the decoding centre and hence crucial for effective mitochondrial translation. The general architecture of the platform region in the human mtSSU is similar to its cytoplasmic and bacterial counterparts, with the 3′ rRNA end binding in the back of the platform (Fig. 3a,b). In the maturing human mtSSU, the 3′ end is stabilized in a sequential, mutually exclusive manner by the two assembly factors ERAL1 and RBFA, both of which are conserved between human mitochondria and bacteria. The discovery of ERAL1 within states A and B enables a direct comparison with bacterial systems in which available structural and biochemical data indicate that the binding modes between mitochondrial ERAL1 and bacterial Era are similar[41]. Biochemical studies of bacterial Era further suggest that this enzyme binds to the pre-ribosome in a GTP-bound state[39,40], a principle that is probably shared with the mitochondrial assembly system.

Fig. 3: rRNA stabilization of the 3′ end and platform formation are coupled in yeast and human assembly.



a, Model of the architecture around the rRNA 3′ end in human state B, illustrating binding of the 3′ end by ERAL1 and pre-recruitment of bS21m. The arrows highlight platform compaction that occurs between states B and C. **b**, Model of the architecture around the rRNA 3′ end in human state C, showing repositioning of the 3′ end via RBFA binding and recruitment of a uS11m–h23 module through bS21m. **c**, Model of the platform architecture in yeast state 1, before backpack and 3′ rRNA incorporation via the backpack. rRNA h45 is not yet incorporated into the particle, and the arrow highlights platform compaction that occurs between states 1 and 2. **d**, Model of the platform architecture in yeast state 2, after backpack docking and stabilization of 3′ rRNA. In parts a–d, conserved protein and rRNA elements share the same colours. **e**, Side view of the platform region rRNA in human state B. The arrows show movement of rRNA elements occurring between states B and C. **f**, Side view of the platform region rRNA in human state C, revealing compaction of the region, h23 binding and accommodation of

NAD. **g**, Side view of the platform region rRNA in yeast state 1. The arrows show movement of rRNA elements occurring between states 1 and 2. **h**, Side view of the platform region rRNA in yeast state 2, revealing compaction of the region and ordering of rRNA h23, h24 and h45.

In states A and B, ERAL1 binds to the rRNA 3′ end with a C-terminal KH domain in a flexible manner, with the GTPase domain binding uS7m. In state B, the presence of ERAL1 precludes docking of h23 and the complete installation of mito r-proteins bS21m and uS11m. As a C-terminal tethering peptide of bS21m is already observed in state B, it is likely that this segment of bS21m is responsible for the initial docking and eventual incorporation of the bS21m–S11m–h23 module in state C upon dissociation of ERAL1 (Fig. 3a,b). A single point mutation in the human ERAL1 G4 motif results in Perrault syndrome, which is characterized by sensorineural deafness and ovarian dysfunction, suggesting that this mutation compromises the ERAL1 GTPase cycle and hence mtSSU platform assembly[29] (Extended Data Fig. 5a–c).

Although ERAL1 binds in a manner allowing for conformational freedom of the rRNA 3′ end, RBFA binds in a stabilizing manner, sandwiching the free rRNA 3′ end between its KH domain, an N-terminal helical peptide of RBFA, and bS21m (Fig. 3a,b). This composite binding site for the rRNA 3′ end observed here is distinct from that observed in *Escherichia coli*, which is formed exclusively by RBFA[23] (Extended Data Fig. 6e). The bS21m-mediated rigidification of the platform observed between states B and C is accompanied by a concerted rRNA compaction that results in the formation of a NAD-binding site at the base of the platform, which stabilizes a series of single-stranded rRNA linkers in the platform (Fig. 3e,f and Extended Data Fig. 7). Together, the replacement of ERAL1 with RBFA, the docking of h23, uS11m and bS21m, and NAD binding aid in the orchestrated compaction towards a rigid platform.

The gradual compaction of the platform has functional consequences for key enzymatic activities. In bacteria, the TFB1M homologue KsgA mediates dimethylation of h45 adenosine residues, an activity that has been shown to be promoted by RBFA[31]. Although we did not observe a bound methyl donor in the enzyme, we did observe methylated target residues in states following dissociation of TFB1M (states D and E) (Extended Data Fig. 3e).

Our reconstructions provide a structural rationale for this biochemical data, suggesting that TFB1M activity is allosterically regulated by the stability of the platform, which is controlled by the 3′ rRNA-binding assembly factors ERAL1 and RBFA. This mechanism is coupled to upstream events in maturation as a NOA1-bound h27 would preclude platform compaction, ensuring that h44 docking occurs before platform formation and h45 methylation. Maturation events downstream of state E subsequently link RBFA and late-stage assembly factor METTL15 function during mitoribosome assembly with mtIF3-mediated translation initiation[17]. Thus, our data describe how NOA1, ERAL1, RBFA and TBF1M act in concert to enforce and coordinate discrete steps in assembly and provide a unified mechanism for platform formation and methylation licensing. Given that all of these factors are shared between human mitochondria and bacteria, this mechanism is probably shared between these assembly pathways.

Key differences in rRNA 3′ end architecture between the human and yeast mtSSU are reflected in the mechanism of 3′ rRNA stabilization during assembly. The yeast mtSSU contains an extended 3′ rRNA, which reaches into the back of the subunit and is bound by a set of yeast-specific mito r-proteins (mS42 and mS43) in a module that we term the backpack (Fig. 3d and Extended Data Fig. 1g). Our biochemical data indicating that the RNA-binding protein Rmd9 also functions as a mitoribosome assembly factor are consistent with the observation that its depletion destabilizes the 15S rRNA[42] (Supplementary Data 2–4). The 3′ rRNA extension in yeast contains a sequence (5′-AAUAUUCUU-3′) that partially matches a motif that is sequence specifically recognized by Rmd9 (ref. [35]). These data suggest that Rmd9 may be involved in yeast 15S rRNA 3′ end maturation in an early and likely conformationally dynamic assembly intermediate. The presence of an extended 3′ rRNA, mS42 and mS43, and the absence of RBFA or ERAL1 in *S. cerevisiae* suggests that stabilization of the 3′ end occurs in a distinct way compared with human and bacterial systems. Consistent with this idea, this backpack module and h45 are flexible and unresolved in state 1, whereas in state 2, these elements have been docked onto the particle (Fig. 3c,d). Docking of the backpack module and 3′ rRNA stabilization results in platform compaction and rRNA ordering similar to that observed in humans (Fig. 3g,h). Given the unique way in which yeast mito r-proteins mS42 and mS43 stabilize the 15S rRNA 3′ end from state 2 onwards, it appears that

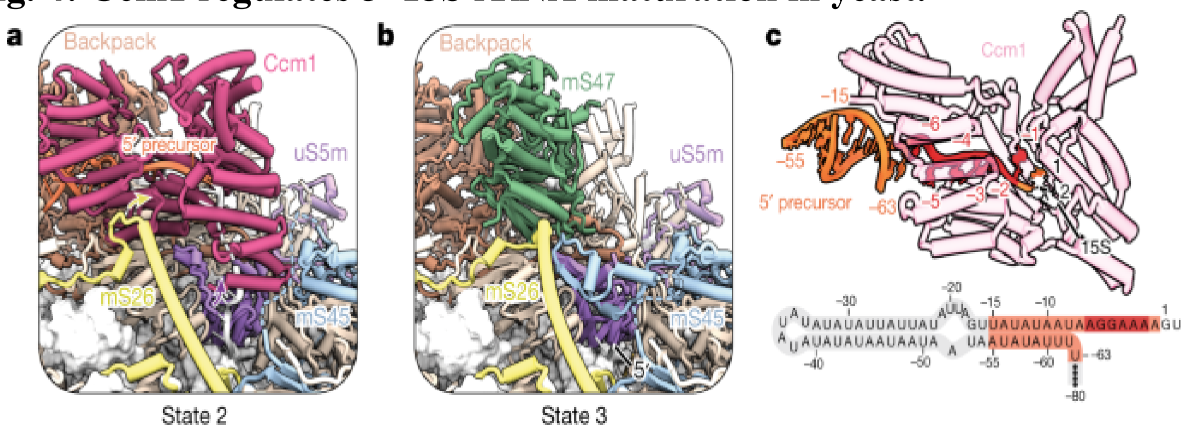
these mito r-proteins have functionally replaced human assembly factors ERAL1 and RBFA. Instead of the controlled 3′ rRNA stabilization mediated by human ERAL1 and RBFA, which is coupled to TFB1M activity, yeast mito r-proteins mS42 and mS43 mediate 3′ rRNA docking and stabilization while remaining bound in the mature mtSSU. Together, these data illustrate the link between 3′ end stabilization and platform maturation and show how rRNA and mito r-protein expansion that occurred through evolution in yeast are accommodated in the yeast assembly system.

Ccm1 controls 15S rRNA 5′ end processing

Unlike human mitoribosomal RNAs, which are punctuated with tRNAs that are post-transcriptionally cleaved and processed to release ‘full-length’ rRNAs, yeast rRNAs must be processed by exonucleases to generate mature subunits[43]. In states 1 and 2, the pentatricopeptide repeat protein Ccm1 localizes to the 5′ precursor of the 15S rRNA in a position that is subsequently occupied by mS47 both in state 3 and in the mature yeast mtSSU (Fig. 4a,b). The binding site and observed function of Ccm1 explain previous biochemical data that implicate this protein in the stabilization of 15S rRNA[33]. Ccm1 uses structure and sequence-specific RNA recognition, binding to a single-stranded region of the precursor and specifically recognizing bases −6 to −1 (Fig. 4c). Ccm1 is therefore in the unique position to chaperone mitochondrial pre-rRNA in a co-transcriptional manner similarly to eukaryotic ribosome assembly factors binding to other 5′ precursor elements[44]. The exchange of Ccm1 for mS47 is coupled to the irreversible removal of precursor rRNA that is accompanied by conformational changes of the mito r-proteins uS5m and mS26m, and additional ordering of a flexible linker of mS45 in state 3 that is previously absent in state 2 (Fig. 4a,b). These conformational changes signal completion of 5′ end rRNA processing through protection of the mature 5′ end of the 15S rRNA and stabilization of mS47 in state 3. The removal of the 5′ precursor together with the dissociation of Ccm1 is at least in part catalysed by the 5′–3′ exoribonuclease Pet127, which has been previously implicated in 15S rRNA and mitochondrial mRNA processing[45]. Consistent with a role in pre-rRNA processing, we found that the deletion of the gene encoding Pet127 results in the accumulation of the 5′ precursor and also affects the association of both translational activators of a small number of

mitochondrial mRNAs as well as mtSSU assembly factors with mtSSUs and their precursors purified via the tagged early binding mito r-protein uS17m (Supplementary Fig. 27 and Supplementary Data 5). Together, our structural and biochemical data show that Ccm1 and Pet127 coordinate the controlled processing of the 5′ end of yeast mitochondrial rRNA.

Fig. 4: Ccm1 regulates 5′ 15S rRNA maturation in yeast.



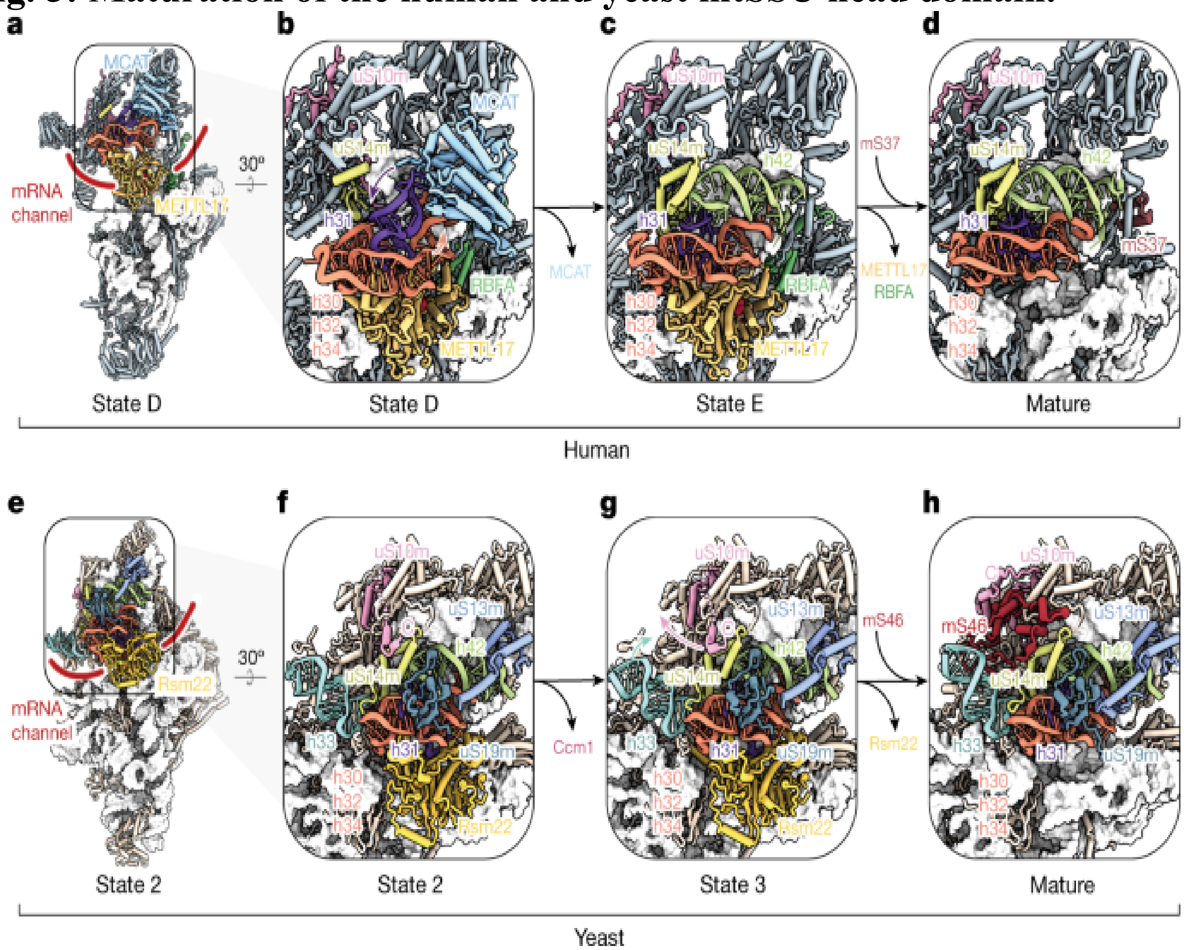
a, Model of Ccm1 and the 5′ precursor in state 2. The arrows indicate conformational changes of mito r-proteins during the transition from state 2 to state 3. **b**, Model of the same region in **a**, but in state 3. **c**, Model of Ccm1 bound to the 5′ precursor rRNA (top) and a secondary structure cartoon of the 5′ precursor sequence (bottom), coloured as in the model: the red region is specifically recognized by Ccm1, the orange regions are resolved, and the grey regions are not resolved.

Related mechanisms of head assembly

The head domain of the ribosomal SSU is responsible for ratcheting along mRNA and tRNA binding during translation. Our reconstructions reveal how assembly factors promote stepwise transformations of rRNA and protein elements in the mtSSU head region. A key conserved assembly factor within mitochondria, which uses steric hindrance during the maturation of the head region, is the putative methyltransferase Rsm22 (METTL17 in human). In both yeast and human systems, Rsm22 (METTL17) binds at the interface of the head and body domains, occluding the mRNA channel and preventing compaction of the head domain towards the body (Fig. 5a,e). The presence of this factor across assembly

intermediates emphasizes its central role as a gatekeeper against premature mRNA binding in earlier stages of assembly, which is coupled to its function in sequestering h31, one of the last elements of the decoding centre to be positioned (Extended Data Fig. 4c–f). We have further identified an Fe_4S_4 cluster as an architectural cofactor of the Rsm22 (METTL17) protein family, which rationalizes previous structural data obtained in kinetoplastids, as well as genetic interactions between METTL17 and iron–sulfur cluster biogenesis machinery[15,46] (Extended Data Fig. 8a–c). The binding of METTL17 to h31 also prevents association of the late-stage assembly factor METTL15, which is only observed in assembly intermediates with a mature head domain[17]. These events enforce a chronology of human mtSSU assembly in which head formation and compaction precede final METTL15-mediated modification of the decoding centre. The architectural role of METTL17 and position in the assembly pathway upstream of METTL15 would further explain rRNA methylation defects observed in METTL17-deficient cells[47].

Fig. 5: Maturation of the human and yeast mtSSU head domain.



a, Model of human state D, highlighting inhibition of mRNA binding via METTL17. **b**–**d**, Top-down zoomed-in view of the head domain across human assembly intermediate state D (**b**), state E (**c**) and the mature pre-initiation structure (PDB: 6RW4) (**d**). **e**, Model of yeast state 2, highlighting inhibition of mRNA binding via Rsm22. **f**–**h**, Top-down zoomed-in view of the head domain across yeast assembly intermediate state 2 (**f**), state 3 (**g**) and the mature structure (PDB: 5MRC) (**h**). Binding and dissociation of protein factors occurring between states are noted between panels. Across all panels, conserved protein and rRNA elements share the same colours, and coloured arrows show conformational changes in RNA or protein elements that occur during assembly.

Our structures show that in humans, METTL17 orchestrates maturation of head rRNA through interplay with an additional assembly factor identified in this study: MCAT (Fig. 5a–d). In states A–D, the presence of both of these factors prevents compaction of the 3′ major rRNA domain, splaying rRNA elements and protein modules apart to allow for significant intermodule flexibility. In addition to sterically hindering folding and docking of h42, which forms a lid along the top of the head domain, MCAT sequesters h31, a primarily single-stranded rRNA segment emanating from h30 and h32, from the core of the head domain (Fig. 5a,b). Upon the departure of MCAT in state E, h42 can become ordered in the position previously occupied by MCAT and h31 can adopt a compact conformation forming contacts with h34, h42 and h43 (Fig. 5c). Thus, dissociation of MCAT triggers global compaction of the 3′ major domain. The identification of MCAT as a bona fide mtSSU assembly factor recontextualizes the observation that mutations in this protein causes LHON-like optic neuropathy[32]. Several mutations in mtDNA encoding subunits of respiratory chain complexes are also associated with this disorder, suggesting that functional defects from these MCAT mutations stem from its role as an mtSSU assembly factor rather than its role in fatty acid metabolism (Extended Data Fig. 5d–f).

Although yeast lack an MCAT homologue, the yeast mtSSU has two additional mito r-proteins present in the same location that MCAT binds in humans, suggesting that they may chaperone h42 docking early in assembly. Although Rsm22 also prevents head compaction in yeast, unlike its homologues in human and *T. brucei* systems, yeast Rsm22 has lost SAM-

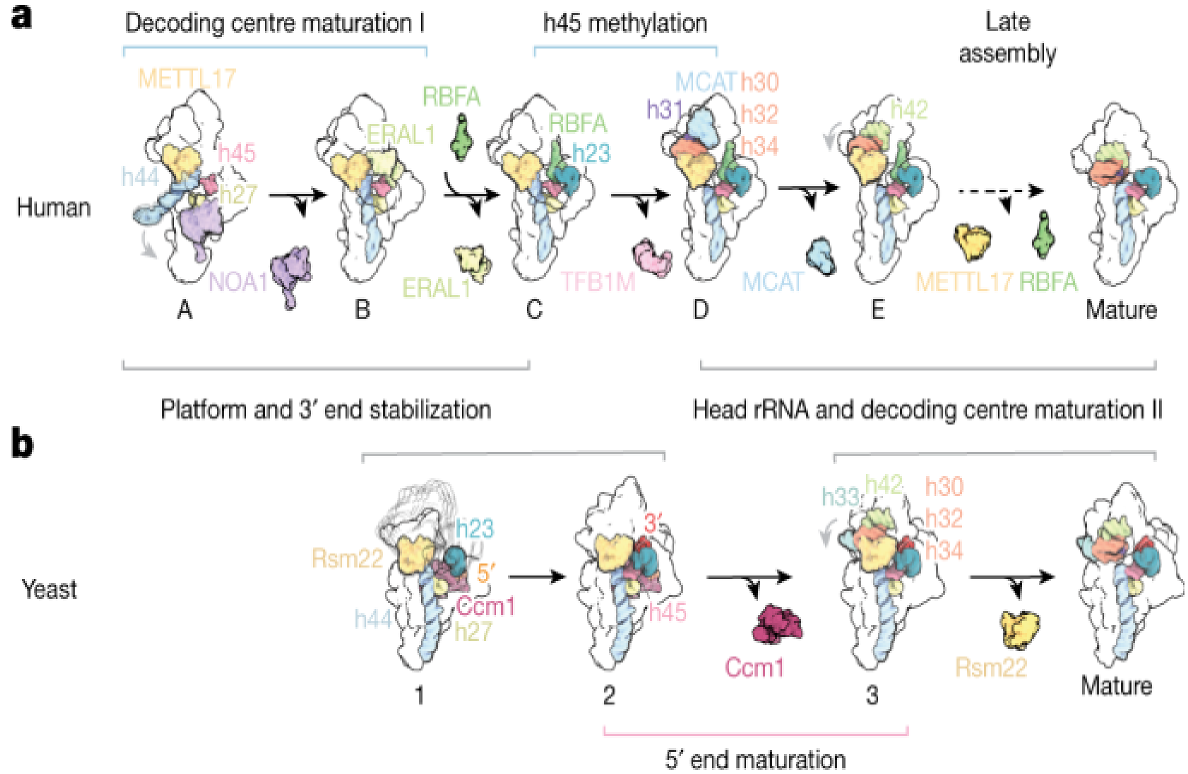
binding capacity but has retained the overall methyltransferase fold (Extended Data Fig. 8a–c). A C-terminal helical peptide segment of uS10m initially prevents the binding of mS46, the final mito r-protein to bind to the head domain. Dissociation of Rsm22 subsequently drives reorientation of h33 and repositioning of this peptide to allow mS46 binding and complete maturation of the head domain (Fig. 5g,h).

In both yeast and human mtSSU assembly systems, maturation of the head domain involves assembly factors acting to sterically hinder discrete rearrangements in rRNA and conserved ordering of mito r-protein elements to promote rRNA packing (Extended Data Fig. 8d–g). In yeast, protein elements such as uS10m seem to have evolved to have an active role during mtSSU assembly by preventing rRNA rearrangements and mito r-protein association while becoming permanent constituents of the mature subunit. This evolutionary trajectory towards dual roles of mito r-proteins in yeast is akin to the adaptations observed for binding, stabilization and incorporation of the 3′ end of the rRNA, in which assembly factors used in human and bacterial systems are exchanged for mito r-proteins binding to an extended rRNA segment.

Discussion

A fundamental challenge during the formation of ribosomal SSUs is the correct assembly of functionally critical rRNA regions such as the decoding centre. Doing so in a controlled and regulated manner is essential for the generation of functional subunits that exhibit high translational fidelity to control the quality of newly synthesized proteins. For mitoribosomes, this aspect of assembly regulation is especially crucial given that its translational products, subunits of oxidative phosphorylation complexes, are essential for creating ATP under aerobic conditions. Here we have solved structures of both human and yeast mitoribosomal SSUs undergoing assembly under native conditions. In addition to uncovering key architectural features of these pre-mitoribosomal subunits and illuminating the roles of assembly factors in regulating the formation of rRNA elements, we have shown how assembly mechanisms and machinery have evolved to solve shared hurdles during assembly between these two model organisms (Fig. 6).

Fig. 6: Shared principles of mitoribosomal SSU assembly in yeast and human cells.



a,**b**, Schematic of human (**a**) and yeast (**b**) mtSSU assembly pathways. The centre brackets indicate common themes across both assembly systems, whereas the outer brackets indicate evolutionarily divergent assembly solutions. Assembly factor binding and dissociation events are shown. Dashed arrows indicate multiple maturation events[17].

In the human assembly system, we have shown that assembly factors in concert with mito r-proteins catalyse stepwise formation and stabilization of rRNA elements before states previously observed[17]. In particular, a network of assembly factors (NOA1, ERAL1, RBFA, TFB1M and METTL17) couple maturation of the decoding centre to platform compaction and 3′ end rRNA stabilization, in the process licensing methylation of key residues in h45 at the correct time during assembly (Fig. 6a). Although the mechanism of 3′ end stabilization in yeast is distinct from that in humans, requiring unique mito r-proteins to bind to and dock an extended rRNA 3′ end rather than dedicated 3′ end binding assembly factors as seen in the human system, we show that stable incorporation of the rRNA 3′ end is a key checkpoint in mtSSU assembly (Fig. 6b). Similarly, yeast uses a distinct 5′ end rRNA

maturation pathway involving a unique assembly factor and yeast-specific mito r-proteins. Conversely, the shared architecture and role of Rsm22 (METTL17) in prevention of mRNA binding, subunit compaction and formation of the decoding centre demonstrates a level of conservation between these two assembly systems, a concept additionally highlighted by similarities in rRNA-folding chronology (Figs. 1i,j and 6). We further note that our study provides a structural view of assembly intermediates that can be associated with Rsm22 (METTL17), thereby visualizing one assembly line of what is probably a branched assembly landscape involving assembly factors dissociating and re-associating at key junctions.

The similarities and differences shown here illustrate how different molecular mechanisms involving dedicated assembly factors and bifunctional mito r-proteins have evolved to solve common hurdles during assembly. These principles highlight evolutionary coupling between assembly mechanism and the final architecture of large, multimeric biomolecular complexes, and shed light on the emergence of molecular complexity and diversity across species.

Methods

HEK293F cell line generation and cell culture

To generate a stable mammalian cell line with endogenously tagged *METTL17* on both alleles, we used a genome editing platform previously developed in the laboratory and used for purification of human pre-ribosomes called SNEAK PEEC (surface engineered fluorescence assisted kit with protein epitope enhanced capture)[24]. This system uses guide RNA-targeted Cas9 genome cleavage and in-frame cell-surface epitope expression to allow for identification of biallelically edited clones by fluorescence-activated cell sorting (FACS). Repair template plasmids contain 600-bp homology arms flanking the Cas9 cut site located in the last exon of *METTL17*, and an in-frame insert containing the affinity tag (linker-twinStrep-FLAG-10xHis-TEV-SPOT-3C-GFP) and one of two engineered cell surface epitopes (the Btuf vitamin B_{12} binding protein from *E. coli* (PDB: 5OVW) and the capsid protein p24 from human immunodeficiency virus 1 (PDB: 5O2U)) separated by a T2A site[48,49].

To generate these repair templates, homology arms flanking the Cas9 target site and a multiple cloning site were cloned into the pUC57 backbone (GenScript). The entire insert sequence, including the affinity tag, T2A site, and the surface display epitopes, were introduced using NotI and PacI sites in the multiple cloning site. The plasmid for expression of single guide RNA and *Streptococcus pyogenes* Cas9 (variant eSpCas9(1.1)) was derived from Addgene plasmid #71814, a gift from F. Zhang[50]. The 20-bp single guide RNA target sequence (5′-GTTTCCTCCATCTACGGCTC-3′, PAM: CGG) was designed using crispor.telfor.net[51] and cloned into this vector using BbsI sites.

HEK293F cells (R79007, Thermo Fisher Scientific) were grown in 24-well plates (353047, Falcon) in Freestyle 293 expression medium (12338026, Thermo Fisher Scientific) supplemented with 2% heat-inactivated fetal bovine serum (FBS). Cell lines were not tested for mycoplasma. Each well was transfected using 1 µg of total DNA (both repair template plasmids and the Cas9/single guide RNA plasmid in equimolar concentrations) at approximately 90% confluency using the transfection reagent Lipofectamine 2000 (11668019, Thermo Fisher Scientific) in 500 µl Opti-MEM medium (31985070, Thermo Fisher Scientific). Twelve hours after transfection, cells were washed and resuspended in Freestyle 293 expression medium supplemented with 2% heat-inactivated FBS. After recovery, cells were expanded to six-well plates (10062-892, VWR) and passaged over the course of 14 days.

To prepare cells for flow cytometry, cells were harvested from six-well plates by gentle aspiration, washed in 1X PBS with 0.1% BSA, and resuspended in 1X PBS with 0.1% BSA to a concentration of $1–10 \times 10^6$ cells per millilitre. Next, two fluorescently labelled nanobodies, each of which specifically binds to one of the epitopes expressed at the cell surface, were added to cell suspensions at a concentration of 10 nM. The labelling reaction was carried out in the dark for 30 min on ice before washing cells twice with 1X PBS with 0.1% BSA. Cells were then resuspended in 200 µl of 1X PBS with 0.1% BSA and filtered directly before cell sorting to remove cell clumps (352235, BD Falcon). Cell sorting was carried out using a BD FACSAria cell sorter (BD Biosciences) using FACSDiva Software (BD Biosciences). DAPI was used to stain dead cells directly before sorting.

Necessary single-colour compensation controls, including GFP and each individual labelled nanobody, were performed directly before sorting of experimental cells using HEK293F cells expressing GFP or cell-surface epitopes from plasmids. To identify biallelically tagged METTL17 cells, cells were first sorted to identify GFP-positive cells, followed by a second selection to isolate cells staining positive for both btuF and p24 surface epitopes. Data analysis was carried out using FlowJo (FlowJo, LLC).

Single-cell clones were sorted into 96-well plates containing 200 µl Freestyle 293 expression medium (12338026, Thermo Fisher Scientific) supplemented with 2% heat-inactivated FBS per well. After 2 weeks of clonal expansion, individual clones were screened by two PCRs to confirm biallelic editing. Genomic DNA was isolated using QuickExtract DNA Extraction Solution (QE09050, Lucigen) according to the manufacturer's protocol. Of extracted solution, 30 µl was used per PCR (50 µl final reaction volume) using the following PCR primers:

For reaction 1, forward: 5′-TATGTGTTCACAGTGTCCCCATGAACTCCC-3′ (genomic primer annealing upstream of the METTL17 left homology arm); reverse: 5′-GACTCCACGGGGCCAACTGTCTCAAGG-3′ (BtuF epitope specific).

For reaction 2, forward: 5′-TATGTGTTCACAGTGTCCCCATGAACTCCC-3′ (genomic primer annealing upstream of the METTL17 left homology arm); reverse: 5′-TGCTGTCATCATTTCCTCGAGCGTAGCACC-3′ (HIV p24 epitope specific).

Biallelically tagged METTL17 HEK293F cells were subsequently grown in suspension in Freestyle 293 expression medium at 37 °C with 80% humidity at 8% CO_2, shaking at 135 r.p.m.

Yeast strains

S. cerevisiae strain BY4741 (MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0) was used as the starting point for construction of endogenously tagged strains. All strains were grown aerobically at 30 °C in YPG media (1% yeast extract, 2% peptone and 3% glycerol). Assembly intermediates of the *S. cerevisiae*

mtSSU were purified from three genetically modified BY4741 strains: the first strain containing a 3C cleavable C-terminal GFP-tagged *MTG3* (MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 *MTG3*-linker-3c-GFP::ClonNat), the second strain containing a TEV protease-cleavable C-terminal mCherry-tagged *CCM1* and a 3C cleavable C-terminal GFP-tagged *MTG3* (MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 *MTG3*-linker-3c-GFP::ClonNat *CCM1*-linker-tev-mCherry::HphMX4), and the third strain containing a C-terminal tandem 3C protease-cleavable mCherry and TEV-cleavable alfa peptide tag on *RSM22* (MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 *RSM22*-linker-tev-alfa tag-3C-mCherry::ClonNat).

For comparison of mitoribosomal complexes in wild-type and *PET127*-knockout (KO) backgrounds, a *PET127*-KO strain was purchased from the Yeast Knockout (YKO) Collection of Horizon Discovery (MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 *PET127*::kanMX6). Wild-type and *PET127*-KO strains were modified to contain a TEV protease-cleavable C-terminal GFP-tagged *US17M* (MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 *US17M*-linker-tev-GFP::HphMX4, MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 *PET127*::kanMX6 *US17M*-linker-tev-GFP::CloNat).

All strains were generated by applying standard genomic tagging techniques. Primers used for yeast genomic tagging are detailed: Ccm1 forward primer:
AAAAGTCACACGCAAAGGCCCTGAAGTGGGAGGAACAAGAACTT
AACATGACGCTGCAGGTCGACGGATCC; Ccm1 reverse primer:
GAGCTTTTATTTACAGGCGCTTTACTATATATATATATGTGCATGCAT
GGGGCAGATCCGCGGCCGCATAGG. Mtg3 forward primer:
TGGTTCAACACCTAGAAAAAAAAAAAGAATCCCTATTGGTACTACC
AATGGACGCTGCAGGTCGACGGATCC; Mtg3 reverse primer:
TGAATATACTTTGTTCCCAACTTTCTTTTTAGTTTACTTACATTCATA
TTGGCAGATCCGCGGCCGCATAGG. Rsm22 forward primer:
AATTGTCCACTTTTCACGGCAACGATTTTTTACAACATGTAAATAG
AAAAACGCTGCAGGTCGACGGATCC; Rsm22 reverse primer:
CAAACTGTACTATATATATATCATATTCACGTATGTAGAATATTAAAG
TAGGCAGATCCGCGGCCGCATAGG. uS17m forward primer:
AAAATTTGTTAAGAAAGCACGTCATGATGGATATGCAACAACCAA
GCCAGACGCTGCAGGTCGACGGATCC; uS17m reverse primer:
TACTATTTACGTAACGAAGTCTTGTAGTTGTATATATTTACAGAGTG
TAGGGCAGATCCGCGGCCGCATAGG.

The described strains were grown aerobically at 30 °C in YPG media (1% yeast extract, 2% peptone and 3% glycerol) until an optical density at 600 nm of 1.5–2. Cells were harvested at 4,000*g* for 5 min, washed once with 1 l of ice cold ddH_2O and once with a volume of ddH_2O supplemented with protease inhibitors (E64, pepstatin and PMSF) equal to the mass of the cell pellet. The final pellet was flash frozen in liquid nitrogen and (if used for complex purification) lysed by cryo-grinding using a Retsch Planetary Ball Mill PM100. Cryo-ground powder was stored at –80 °C or used immediately.

Purification of human pre-mitoribosomal complexes

Biallelically tagged METTL17 HEK293F cells were grown in suspension in Freestyle 293 expression medium at 37 °C with 80% humidity and 8% CO_2, shaking at 135 r.p.m. Of cells, 3.2 l were grown to 3.0×10^6 cells per millilitre and harvested by centrifugation at 2,000*g* (JLA 8.100, Beckman Coulter). Cells were washed in ice-cold 1X PBS, dropped into liquid N_2 and frozen at −80 °C until use. Crude mitochondrial purification was completed using a protocol adapted from ref. [52]. All steps were completed on ice using pre-cooled buffers at 4 °C, and all centrifugation steps were completed at 4 °C. Frozen cells were resuspended in ice-cold hypotonic buffer (20 mM HEPES-KOH pH 7.5, 5 mM KCl, 1.5 mM $MgCl_2$ and with 1X protease inhibitors (E64, pepstatin and PMSF)) at a ratio of 1 ml g^{-1} of frozen material. Cells swelled for 10 min before the resuspension was transferred to a dounce homogenizer. Cells were then lysed by 15 strokes of pestle B. 2.5X MSH buffer (525 mM mannitol, 175 mM sucrose, 20 mM HEPES pH 7.5 with 1X protease inhibitors (E64, pepstatin and PMSF)) was added to 1X concentration. Lysate was then cleared by two rounds of centrifugation at 1,300*g* for 5 min to sediment cell debris. Supernatant was then spun at 15,000*g* for 15 min (JA 25.50) to pellet mitochondria. Crude mitochondria were washed with 1X MSH buffer and resuspended in mitochondrial resuspension buffer (25 mM HEPES-KOH pH 7.5, 100 mM KCl, 25 mM $Mg(OAc)_2$ and 1% Triton X-100 v/v with 1X protease inhibitors (E64, pepstatin and PMSF)) at a ratio of 1 ml per 2.5 g frozen material. Mitochondria were lysed for 30 min at 4 °C with constant mixing, and lysate was cleared by centrifugation at 17,000*g* for 20 min (JA 25.50, Beckman

Coulter). Of pre-equilibrated anti-GFP beads (Chromotek), 30 µl was added to the supernatant and incubated for 3 h at 4 °C with constant mixing. Beads were washed twice with mitochondrial resuspension buffer, and once with wash buffer (25 mM HEPES-KOH pH 7.5, 100 mM KCl and 25 mM $Mg(OAc)_2$ with 1X protease inhibitors (E64, pepstatin and PMSF)). Beads were then resuspended in 40 µl elution buffer (25 mM HEPES-KOH pH 7.5, 100 mM KCl, 25 mM $Mg(OAc)_2$ and 0.005% C12E8), and eluted by addition of 3C protease over 1 h on ice. Eluted complexes were used immediately for grid preparation, with analysis by LC–MS or SDS–PAGE.

Cryo-EM grid preparation and data acquisition of human complexes

Grids were prepared using a Vitrobot Mark IV robot (FEI Company) at 95% humidity. Of eluted complex solution, 3.5 µl was applied to a glow discharged Quantifoil Au R3.5/1 with a layer of 2-nm ultrathin carbon (LFH7100AR35, Electron Microscopy Sciences). The grid was incubated for 2.5 min inside the Vitrobot chamber before manually blotting excess solution. Of fresh sample, 3.5 µl was then reapplied before another 2.5-min incubation and manual blotting. A final 3.5-µl sample application and 2.5-min incubation was completed before blotting with a blot force of 8, a blot time of 8 and plunging into liquid ethane. Grids were imaged on a Titan Krios electron microscope (FEI) with an energy filter (slit width of 20 eV) and a K3 Summit detector (Gatan) operating at 300 kV with a nominal magnification of ×64,000. SerialEM[53] was used to collect a dataset totalling 37,047 micrographs with a defocus range of −0.7 to −1.8 µm and a super-resolution pixel size of 0.54 Å. Micrographs contained 42 frames using a total dose of 28.5 e^{-} per pixel per second with an exposure time of 2.1 s, resulting in a total dose of 51 e^{-} $Å^{-2}$. Using a multi-shot strategy, nine micrographs were collected per hole, and four holes were imaged at each stage position. A second dataset comprising 9,990 micrographs was collected on the same grid on different holes using a different collection strategy in which nine holes were imaged at each stage position, using the same defocus range, electron dose and frame count. These datasets are denoted dataset 1 (37,047 micrographs) and dataset 2 (9,990 micrographs) in the supplementary figures and below.

Cryo-EM data processing of human complexes

All cryo-EM processing steps were completed using RELION 3.1.1 (ref. [54]) unless noted, and are summarized in Supplementary Figs. 3–18. For each dataset, images were gain corrected, dose weighted, aligned and binned to a pixel size of 1.08 Å using MotionCor2 implementation in RELION[55], and micrograph defocus was estimated using GCTF1.18 (ref. [56]). An initial model for 3D classification was generated by using a test subset of 2,597 micrographs from dataset 1 as follows: CrYOLO[57] was trained and used to pick 533,056 particles from these micrographs, and these particles were extracted at 8.64 Å per pixel (8× binning). After 2D classification ($K = 100$ and $T = 2$), six classes were conservatively picked, which resembled different views of the mature mtSSU, resulting in 65,174 particles, which were subsequently re-extracted at a pixel size of 4.32 Å per pixel (4× binning). These particles were used to generate an ab initio initial model. This initial model was used as a reference for 3D classification with global alignment of a second set of 333,265 particles from the 2D classification, which were picked less conservatively and re-extracted at 2.16 Å per pixel (2× binning). Two resulting classes, comprising 80% of the data and 268,009 particles, were combined and refined to 4.36 Å. This reconstruction was used as a reference for classification of the entire dataset.

CrYOLO was trained independently on each dataset and used to pick particles from all micrographs, resulting in 7,597,700 particles for dataset 1 and 1,511,635 particles for dataset 2. These particles were extracted at 8.64 Å per pixel (8× binning), and dataset 1 was split into four subsets with 1,894,925 particles each. To clean the particle stack, each subset, as well as all particles from dataset 2, were subjected to 2D classification ($K = 100$ and $T = 2$) followed by 3D classification with alignment ($K = 5$ and $T = 4$; 25 iterations with 7.5° angular sampling, 10 iterations with 3.7° angular sampling and 10 iterations with 1.8° angular sampling). Good classes from each 3D classification were combined and duplicates were removed, resulting in 1,927,789 total particles, with particles from each dataset having different optics groups. Contrary to the test dataset, the selected particles comprised approximately only 40% of particles from each subset. To ensure all good particles were selected, all particles from initially unselected classes were combined and re-extracted at 4.32 Å per pixel (4× binning), totalling

4,015,406 particles. These particles were split into two subsets of 2,007,700 particles each and subjected to a second round of 3D classification with alignment ($K = 5$ and $T = 4$; 25 iterations with 7.5° angular sampling, 10 iterations with 3.7° angular sampling and 10 iterations with 1.8° angular sampling). Good classes were combined and duplicates were removed, resulting in 1,183,364 particles, which were merged with the initial selection of 1,927,789 particles. After removing duplicates, a final particle stack with 2,575,411 particles was re-extracted at 1.08 Å per pixel (1× binning), refined and subjected to Bayesian polishing and contrast transfer function (CTF) refinement. A subsequent refinement resulted in a 2.45 Å consensus map.

To determine the compositional and conformational heterogeneity present in the consensus map, 3D classification without alignment was used. Initial attempts were done without a mask, which resulted in classes with different relative orientations between the head and body domains, reflecting the conformational freedom of the head domain. However, this strategy was not optimal for distinguishing compositional differences in the data, especially those present in the body, platform and head regions. As a result, 3D classification without alignment ($K = 10$ and $T = 30$) was performed with a mask around the body domain, which successfully resolved the compositional heterogeneity present. Four classes were identified, which represented compositionally unique assembly intermediates. Of the data, 10% (263,184 particles) represent the earliest state in which h44 is displaced, the density for NOA1 and TFB1M are present and the platform is less ordered. We denote this intermediate state A. Of the data, 19% (497,031 particles) represent the next state, in which h44 is present and NOA1 is absent, but the platform is still remodelled and TFB1M is present, as in state A. We denote this intermediate state B. Of the data, 10% (262,363 particles) represent a state in which the platform is well ordered and TFB1M is present. We denote this intermediate state C. Of the data, 27% (697,003 particles) represent a state in which TFB1M is not present and the platform is well ordered. We denote this intermediate state D. Each of these states were subjected to another round of CTF refinement and Bayesian polishing, refined and subsequently processed independently.

To fully structurally characterize these intermediates, we adopted a strategy using focused classification and/or focused refinement to improve local resolutions of flexible regions, such as those occupied by assembly factors,

the head and the platform. In particular, the head domain exhibits significant conformational freedom relative to the body domain, as is seen in other ribosomal SSU reconstructions. For each state, focused refinement was used to realign the particles on the head domain, which revealed significant remodelling of rRNA due to the presence of the assembly factors METTL17 and MCAT. The resulting modules exhibit additional relative flexibility, further limiting resolution of these regions. To address this, the head domain for each state was subjected to focused 3D classification without alignment to identify a subset of particles with well-defined modules before focused classification and/or focused refinement of individual modules. For states C and D, this classification resulted in the identification of small subsets of data in which MCAT is absent and h42 is stabilized in a near-mature conformation (Supplementary Figs. 6c and 7d). We denote these states state C* and state E, respectively, and further processed these particles independently (Supplementary Figs. 8 and 9). We propose that state C* represents an intermediate of a secondary pathway in which MCAT dissociation occurs before dissociation of TFB1M (thus skipping state D), whereas state E represents an intermediate directly following state D and state C*. All focused refinements were performed using solvent flattening, and reported resolutions are based on the gold-standard FSC (Fourier Shell Correlation)-0.143 criterion as calculated during post-processing in RELION. Local resolutions of focused maps were calculated in RELION.

To provide a complete view of the assembly process, we exploited overlap in body and head maps to generate composite maps and corresponding models that represent the predominant relative conformation between these domains. Focused maps were fit and resampled into the overall maps in Chimera[58] and were subsequently combined to create composite maps using phenix.combine_focused_maps[59]. Half-maps from each focused refinement were also combined using the same procedure. The composites were then rescaled to a calibrated pixel size of 1.062 Å per pixel, and overall resolutions were calculated in RELION. cryoSPARC[60] was used to calculate local resolution of composite maps and to generate maps filtered by local resolution. The Remote 3D-FSC Processing Server was used to calculate 3D-FSC curves for composite maps[61].

Model building and refinement of human complexes

For human states, a starting model including all ribosomal proteins and 12S rRNA (PDB: 6RW4) was used as an initial template for model building by rigid body docking into composite maps. To build assembly factors, a combination of homology models generated in the AlphaFold database (NOA1, METTL17, RBFA and ERAL1) and existing X-ray structures (MCAT (PDB: 2C2N) and TFB1M (PDB: 6AAX)) were used as starting models before manual adjustment using COOT[62]. Changes in RNA conformation were accounted for in each state by manual refinement in COOT[62]. Finally, entire models for each state were adjusted with three cycles of refinement in PHENIX using phenix.real_space_refine using secondary structure restraints[59]. Owing to lower local resolution for the ERAL1 KH domain, we trimmed side chains in this section of the chain to C_β after all-atom refinement. Model refinement statistics for human models can be found in Extended Data Table 1.

Maps and models were visualized in Chimera[58], ChimeraX[63] and PyMol (Schrödinger, LLC).

Figures were generated using ChimeraX[63]. PyMol sessions are available for each state individually and with all states together.

Purification of Ccm1-associated mtSSU intermediates

Cryo-ground yeast powder was resuspended in buffer A (25 mM HEPES-KOH pH 7.5, 100 mM KCl, 25 mM MgOAc, 0.1% Triton X-100, 2 mM DTT, PMSF, pepstatin and E64) and cleared by centrifugation at 4 °C and 40,000*g* for 20 min. The cleared lysate was incubated with NHS–sepharose beads (Cytiva) coupled to anti-mCherry nanobodies for 3 h at 4 °C on a nutator. Beads were pelleted by centrifugation at 4 °C for 1 min at 127*g*. After three washes in buffer A, complexes were eluted in buffer B (25 mM HEPES-KOH pH 7.5, 100 mM KCl and 25 mM MgOAc) supplemented with TEV protease for 1 h at 4 °C.

Mtg3-associated mtSSU intermediate purification

Cryo-ground yeast powder was resuspended in buffer A (25 mM HEPES-KOH pH 7.5, 100 mM KCl, 25 mM MgOAc, 0.1% Triton X-100, 2 mM

DTT, PMSF, pepstatin and E64) and cleared by centrifugation at 4 °C and 40,000*g* for 20 min. The cleared lysate was incubated with anti-GFP nanobody beads (Chromotek) for 3 h at 4 °C on a nutator. Beads were pelleted by centrifugation at 4 °C for 1 min at 127*g*. After three washes in buffer A, complexes were eluted in buffer B (25 mM HEPES-KOH pH 7.5, 100 mM KCl and 25 mM MgOAc) supplemented with 3C protease for 1 h at 4 °C.

Purification of Rsm22-associated mtSSU intermediates

Cryo-ground yeast powder was resuspended in buffer A (25 mM HEPES-KOH pH 7.5, 100 mM KCl, 25 mM MgOAc, 0.1% Triton X-100, 2 mM DTT, PMSF, pepstatin and E64) and cleared by centrifugation at 4 °C and 40,000*g* for 20 min. The cleared lysate was incubated with NHS–sepharose beads (Cytiva) coupled to anti-mCherry nanobodies for 3 h at 4 °C on a nutator. Beads were pelleted by centrifugation at 4 °C for 1 min at 127*g*. After three washes in buffer A, the immobilized complexes were incubated with 3C protease at 4 °C for 1 h. After protease cleavage, the supernatant was applied to NHS–sepharose beads (Cytiva) coupled to anti-alfa nanobodies and incubated in buffer B (25 mM HEPES-KOH pH 7.5, 100 mM KCl and 25 mM MgOAc) for 1.5 h at 4 °C. Beads were subsequently washed three times in buffer B and the complex was eluted in the same buffer, supplemented with TEV protease.

Purification of uS17m-associated RNA–protein complex from *PET127*-KO and wild-type yeast cells

Cryo-ground yeast powder was resuspended in buffer A (25 mM HEPES-KOH pH 7.5, 100 mM KCl, 25 mM MgOAc, 0.1% Triton X-100, 2 mM DTT, PMSF, pepstatin and E64) and cleared by centrifugation at 4 °C and 40,000*g* for 20 min. The cleared lysate was incubated with anti-GFP nanobody beads (Chromotek) for 3 h at 4 °C on a nutator. Beads were pelleted by centrifugation at 4 °C for 1 min at 127*g*. After three washes in buffer A, complexes were eluted in buffer B (25 mM HEPES-KOH pH 7.5, 100 mM KCl and 25 mM MgOAc) supplemented with TEV protease for 1 h at 4 °C.

Cryo-EM grid preparation and data collection for yeast particles

Grids were prepared using a Vitrobot Mark IV robot (FEI Company) at 100% humidity. For the Ccm1 dataset, 3.5 µl eluted complex solution was applied to a glow-discharged Quantifoil Au R3.5/1 with a layer of 2-nm ultrathin carbon (LFH7100AR35, Electron Microscopy Sciences). The grid was incubated for 1 min inside the Vitrobot chamber before manually blotting excess solution. Of fresh sample, 3.5 µl was then reapplied before another 1-min incubation and manual blotting. A final 3.5 µl sample application and 1-min incubation were completed before adding CHAPSO to CMC (8 mM). The grid was then blotted (blot force of 8 and blot time of 8 s) and plunged into liquid ethane.

For the Rsm22 dataset, 3.5 µl eluted complex solution was applied to a glow-discharged Quantifoil Au R3.5/1 grid with a layer of 2-nm ultrathin carbon (LFH7100AR35, Electron Microscopy Sciences). The grid was incubated for 2.5 min inside the Vitrobot chamber before adding CHAPSO to CMC (8 mM). The grid was then blotted (blot force of 8 and blot time of 3.5 s) and plunged into liquid ethane.

Grids were imaged on a Titan Krios electron microscope (FEI) with an energy filter (slit width of 20 eV) and a K3 Summit detector (Gatan) operating at 300 kV with a nominal magnification of ×64,000. SerialEM[53] was used to collect two datasets totalling 31,995 (Ccm1) and 14,111 (Rsm22) micrographs with a defocus range of −1 to −2.5 µm and a super-resolution pixel size of 0.54 Å. Micrographs contained 40 frames using a total dose of 36 e^{-}per pixel per second (specimen pixel size of 1.08 Å per pixel) with an exposure time of 2 s and a total dose of 61.73 e^{-} $Å^{-2}$. A different multi-shot strategy was used for each dataset. For the Ccm1 dataset, a multi-shot strategy was used with nine micrographs collected per hole, and nine holes were imaged at each single stage position. For the Rsm22 dataset, ten micrographs were collected per hole, and four holes were imaged at each stage position.

Cryo-EM data processing of yeast particles

All cryo-EM processing steps were completed using RELION 3.1.1 (ref. [54]) unless noted, and are summarized in Supplementary Figs. 21–25. For each dataset, images were gain corrected, dose weighted, aligned and binned to a pixel size of 1.08 Å using the MotionCor2 implementation in RELION[55], and micrograph defocus was estimated using GCTF1.18 (ref. [56]).

For the Ccm1-associated mtSSU intermediates, an initial model for 3D classification was generated by using a test subset of 8,189 micrographs as follows: CrYOLO[57] was trained and used to pick 830,061 particles from these micrographs, and these particles were extracted at 4.32 Å per pixel (4× binning). After two rounds of 2D classification ($K = 100$ and $T = 2$), four classes were conservatively picked, which resembled different views of the mature mtSSU, resulting in 23,206 particles. These particles went through two rounds of ab initio reconstruction in cryoSPARC[60]. Refinement of 7,263 selected particles in cryoSPARC allowed generation of an ab initio initial model.

CrYOLO was trained and used to pick particles from all micrographs, resulting in 3,286,640 particles. These particles were extracted at 4.32 Å per pixel (4× binning) and split into 5 subsets with 357,328 particles each. To clean the particle stack, each subset was subjected to 3D classification with alignment ($K = 5$ and $T = 4$; 25 iterations with 7.5° angular sampling, 10 iterations with 3.7° angular sampling and 10 iterations with 1.8° angular sampling). Good classes from each 3D classification were combined and duplicates were removed, resulting in 316,817 total particles. These particles were re-extracted at 1.08 Å per pixel (1× binning), refined and subjected to Bayesian polishing and CTF refinement. A subsequent refinement resulted in a 3.13 Å consensus map. To identify conformationally distinct states in the data, two rounds of 3D classification without alignment ($K = 3$ and $T = 50$) were performed with a mask around the body domain. Two distinct classes were identified, which represented compositionally unique assembly intermediates. Both states contain a 5′ precursor bound to the mtSSU assembly factor Ccm1. 19,730 particles represent the earliest state in which the platform is disordered and a group of ribosomal proteins—which we call the backpack (bS1m, uS2m, mS23, mS37, mS42 and mS43)—that bind to the 3′ end of the 15S rRNA are absent. We refer to this intermediate as state 1. 54,519 particles represent a later state in which the platform is well

ordered and backpack proteins are bound. We denote this intermediate as state 2. Each of these states were refined and subsequently processed independently.

To fully structurally characterize these intermediates, we adopted a strategy using focused classification and/or focused refinement to improve local resolutions of flexible regions, such as those occupied by assembly factors, the head and the platform. All focused refinements were performed using solvent flattening, and reported resolutions are based on the gold-standard FSC-0.143 criterion as calculated during post-processing in RELION. The local resolution of focused maps was calculated in RELION.

For the Rsm22-associated mtSSU intermediates, an initial model for 3D classification was generated by using a test subset of 3,982 micrographs as follows: CrYOLO was trained and used to pick 891,434 particles from these micrographs, and these particles were extracted at 4.32 Å per pixel (4× binning). After 2D classification ($K = 100$ and $T = 2$), 17 classes were picked, which resembled different views of the mature mtSSU, resulting in 635,817 particles. These particles went through ab initio reconstruction in cryoSPARC allowing generation of an ab initio initial model.

CrYOLO was trained and used to pick particles from all micrographs, resulting in 3,544,843 particles, which were extracted at 8.64 Å per pixel (8× binning). To clean the particle stack, all particles were subjected to 2D classification ($K = 100$ and $T = 2$) followed by 3D classification with alignment ($K = 5$ and $T = 4$; 20 iterations with 7.5° angular sampling, 10 iterations with 3.7° angular sampling and 10 iterations with 1.8° angular sampling). Good classes from this classification were combined resulting in 1,656,825 total particles. These particles were re-extracted at 1.08 Å per pixel (1× binning), refined and subjected to Bayesian polishing and CTF refinement. A subsequent refinement resulted in a 2.84 Å consensus map. It was clear from the consensus map that there was considerable flexibility in the head domain of this particle subset. To visualize this region, 3D classification without alignment ($K = 3$ and $T = 50$) was performed with a mask around the head domain. Two good classes from this state were combined and subjected to Bayesian polishing, CTF refinement and 3D refinement, resulting in a 2.65 Å map. To fully structurally characterize this intermediate, we used multibody refinement with masks around the head and

body domains, resulting in two separate maps. This improved the local resolution, most notably in the head domain. Reported resolutions of these maps are based on the gold-standard FSC-0.143 criterion as calculated during post-processing in RELION. The local resolution of focused maps were calculated in RELION.

We exploited overlap in the body and head maps to generate composite maps and corresponding models that represent the predominant relative conformation between these domains for each individual state. Focused maps were fit and resampled into the overall maps in Chimera[58] and subsequently combined to create composite maps using phenix.combine_focused_maps[59]. Half-maps from each focused refinement or multibody refinement were also combined using the same procedure. The composites were then rescaled to a calibrated pixel size of 1.057 Å per pixel, and overall resolutions were calculated in RELION. cryoSPARC was used to calculate local resolution of composite maps and to generate maps filtered by local resolution. The Remote 3D-FSC Processing Server was used to calculate 3D-FSC curves for composite maps[61].

Model building and refinement for yeast structures

For yeast states, a starting model including all ribosomal proteins and 15S rRNA (PDB: 5MRC) was used as an initial template for model building by rigid body docking into composite maps. To build assembly factors, homology models of Rsm22 and Ccm1 generated using Alphafold[64] were used as starting models before manual adjustment using COOT[62]. Changes in RNA conformation were accounted for in each state by manual refinement in COOT[62]. Finally, entire models for each state were adjusted with three cycles of refinement in PHENIX using phenix.real_space_refine using secondary structure restraints[59]. Owing to lower local resolution of Rsm22 in state 1, we trimmed side chains of this protein to C_β after all-atom refinement, excluding residues essential for iron–sulfur cluster coordination. Model refinement statistics for yeast models can be found in Extended Data Table 2.

Maps and models were visualized in Chimera[58], ChimeraX[63] and PyMol (Schrödinger, LLC). Figures were generated using ChimeraX[63]. PyMol

sessions are available for each state individually and with all states together.

RNA extraction and northern blotting of yeast mitoribosomal RNA

Total cellular RNA was extracted from 0.2 g of frozen yeast cells after lysis by bead beating in 1 ml TRIzol (Ambion). For the analysis of pre-rRNA processing states by northern blotting, 3 µg total cellular RNA was loaded in each lane of a denaturing 1.2% formaldehyde–agarose gel (SeaKem LE, Lonza) and separated at 75 V for 3 h. After running, the separated RNA was transferred onto a cationized nylon membrane (Zeta-Probe GT, Bio-Rad) using downward capillary transfer and crosslinked to the membrane for northern blot analysis by UV irradiation at 254 nm with a total exposure of 120 mJ cm^{-2} in a UV Stratalinker 2400 (Stratagene).

Before the addition of γ-32P-end-labelled DNA oligo nucleotide probes, crosslinked membranes were incubated with hybridization buffer (750 mM NaCl, 75 mM trisodium citrate, 1% (w/v) SDS, 10% (w/v) dextran sulfate and 25% (v/v) formamide) for 30 min at 65 °C. Labelled hybridization probes were incubated with the membrane first at 65 °C for 1 h and then at 37–45 °C overnight. Blotted membranes were washed once with wash buffer 1 (300 mM NaCl, 30 mM trisodium citrate and 1% (w/v) SDS) and once with wash buffer 2 (30 mM NaCl, 3 mM trisodium citrate and 1% (w/v) SDS) for 30 min each at 45 °C, before the radioactive signal was read out by exposure of the washed membranes to a storage phosphor screen, which was subsequently scanned with a Typhoon 9400 variable-mode imager (GE Healthcare). Oligonucleotide probe sequences were as follows: 15S, GTTTACTACTAGAACTACACGGGTATCG; 5′ precursor, ATTTTTTACTTTTCCTTATTATA.

MS analysis

Purified ribonucleoprotein samples were dried and dissolved in 8 M urea, 0.1 M ammonium bicarbonate and 10 mM DTT. After reduction, cysteines were alkylated in 30 mM iodoacetamide (Sigma). Proteins were digested with LysC (endoproteinase LysC, Wako Chemicals) in less than 4 M urea followed by trypsination (trypsin gold, Promega) in less than 2 M urea.

Digestions were halted by adding TFA, and digests were desalted[65] and analysed by reversed phase nano-LC–MS/MS using a Fusion Lumos (Thermo Scientific). Data were quantified and searched against the *S. cerevisiae* or *H. sapiens* Uniprot protein database (2019) concatenated with the MS2 protein sequence and common contaminations. For the search and quantitation, MaxQuant v2.0.3.0 (ref. [66]) was used. Oxidation of methionine and protein N-terminal acetylation were allowed as variable modifications and all cysteines were treated as being carbamidomethylated. The 'match between runs' option was enabled, and false discovery rates for proteins and peptides were set to 1% and 2%, respectively.

For comparative MS, protein abundances were expressed as label-free quantitation values. Data were analysed using Perseus (v1.6.10.50)[67]. In short, label-free quantitation values were $\log_2$ transformed followed by a filtering requiring that a protein must be matched in all three replicates for at least one condition. A two-sided Student's *t*-test was carried out to confirm statistical significance of data.

All MS data are available in Supplementary Data 1–5.

Western blotting

To determine whether mitochondrial translation is compromised in METTL17–GFP-tagged cells, wild-type and METTL17–GFP HEK293F cell pellets were resuspended in lysis buffer (25 mM HEPES-KOH pH 7.5, 100 mM KCl, 25 mM $Mg(OAc)_2$ and 1% Triton X-100, with 2X protease inhibitors (E64, pepstatin and PMSF)) and lysed for 30 min at 4 °C. Lysate was cleared at 17,000*g* for 20 min at 4 °C, and total protein concentration was determined by the Bio-Rad Protein Assay Kit. Equal amounts of total protein (20 µg) was separated by SDS–PAGE and transferred to a polyvinylidene difluoride membrane before blocking in TBS-T with 5% skimmed milk for 1 h at room temperature. The membrane was then incubated with primary antibody in TBS-T with 3% skimmed milk for 2 h at room temperature at the following dilutions: MT-CO1 (clone 1D6E1A8, cat. #459600, Thermo Fisher Scientific) for 1:1,000, MT-ND1 (clone 5J5C8, cat. #MA5-42939, Thermo Fisher Scientific) for 1:1,000 and β-actin (cat. #PA1-183, Thermo Fisher Scientific) for 1:2,000. The membrane was then washed

three times in TBS-T before incubation with the corresponding horseradish peroxidase (HRP)-conjugated secondary antibody: HRP-conjugated goat anti-rabbit IgG (cat. #111-035-003, Jackson ImmunoResearch) or HRP-conjugated goat anti-mouse IgG (cat. #115-035-003, Jackson ImmunoResearch) at a dilution of 1:5,000 in TBS-T with 3% skimmed milk. The membrane was then washed twice in TBS-T and once in TBS before imaging on a Typhoon 9400 imager (GE) using ECL Prime detection reagent (Amersham).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Raw cryo-EM micrographs have been deposited in the Electron Microscopy Public Image Archive: EMPIAR-11313. The cryo-EM maps and atomic models have been deposited in the Electron Microscopy Data Bank (EMDB) and the PDB, respectively: state A (EMD-26966 and 8CSP), state B (EMD-26967 and 8CSQ), state C (EMD-26968 and 8CSR), state D (EMD-26969 and 8CSS), state E (EMD-26970 and 8CST), state C* (EMD-26971 and 8CSU), state 1 (EMD-27249 and 8D8J), state 2 (EMD-27250 and 8D8K) and state 3 (EMD-27251 and 8D8L).

References

1. Lopez Sanchez, M. I. G., Krüger, A., Shiriaev, D. I., Liu, Y. & Rorbach, J. Human mitoribosome biogenesis and its emerging links to disease. *Int. J. Mol. Sci.* **22**, 3827 (2021).

2. Tobiasson, V. & Amunts, A. Ciliate mitoribosome illuminates evolutionary steps of mitochondrial translation. *eLife* **9**, e59264 (2020).

3. Ramrath, D. J. F. et al. Evolutionary shift toward protein-based architecture in trypanosomal mitochondrial ribosomes. *Science* **362**, eaau7735 (2018).

4. Soufari, H. et al. Structure of the mature kinetoplastids mitoribosome and insights into its large subunit biogenesis. *Proc. Natl Acad. Sci. USA* **117**, 29851–29861 (2020).

5. Waltz, F., Soufari, H., Bochler, A., Giegé, P. & Hashem, Y. Cryo-EM structure of the RNA-rich plant mitochondrial ribosome. *Nat. Plants* **6**, 377–383 (2020).

6. Waltz, F. et al. How to build a ribosome from RNA fragments in *Chlamydomonas* mitochondria. *Nat. Commun.* **12**, 7176 (2021).

7. Jaskolowski, M. et al. Structural insights into the mechanism of mitoribosomal large subunit biogenesis. *Mol. Cell* **79**, 629–644.e4 (2020).

8. Tobiasson, V. et al. Interconnected assembly factors regulate the biogenesis of mitoribosomal large subunit. *EMBO J.* **40**, e106292 (2021).

9. Cipullo, M., Gesé, G. V., Khawaja, A., Hällberg, B. M. & Rorbach, J. Structural basis for late maturation steps of the human mitoribosomal large subunit. *Nat. Commun.* **12**, 3673 (2021).

10. Hillen, H. S. et al. Structural basis of GTPase-mediated mitochondrial ribosome biogenesis and recycling. *Nat. Commun.* **12**, 3672 (2021).

11. Lenarčič, T. et al. Stepwise maturation of the peptidyl transferase region of human mitoribosomes. *Nat. Commun.* **12**, 3671 (2021).

12. Cheng, J., Berninghausen, O. & Beckmann, R. A distinct assembly pathway of the human 39S late pre-mitoribosome. *Nat. Commun.* **12**, 4544 (2021).

13. Chandrasekaran, V. et al. Visualizing formation of the active site in the mitochondrial ribosome. *eLife* **10**, e68806 (2021).

14. Rebelo-Guiomar, P. et al. A late-stage assembly checkpoint of the human mitochondrial ribosome large subunit. *Nat. Commun.* **13**, 929 (2022).

15. Saurer, M. et al. Mitoribosomal small subunit biogenesis in trypanosomes involves an extensive assembly machinery. *Science* **365**, 1144–1149 (2019).

16. Lenarčič, T. et al. Mitoribosomal small subunit maturation involves formation of initiation-like complexes. *Proc. Natl Acad. Sci. USA* **119**, e2114710118 (2022).

17. Itoh, Y. et al. Mechanism of mitoribosomal small subunit biogenesis and preinitiation. *Nature* **606**, 603–608 (2022).

18. Amunts, A. et al. Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343**, 1485–1489 (2014).

19. Desai, N., Brown, A., Amunts, A. & Ramakrishnan, V. The structure of the yeast mitochondrial ribosome. *Science* **355**, 528–531 (2017).

20. Brown, A. et al. Structure of the large ribosomal subunit from human mitochondria. *Science* **346**, 718–722 (2014).

21. Amunts, A., Brown, A., Toots, J., Scheres, S. H. W. & Ramakrishnan, V. Ribosome. The structure of the human mitochondrial ribosome. *Science* **348**, 95–98 (2015).

22. Vanden Broeck, A. & Klinge, S. An emerging mechanism for the maturation of the small subunit processome. *Curr. Opin. Struct. Biol.* **73**, 102331 (2022).

23. Schedlbauer, A. et al. A conserved rRNA switch is central to decoding site maturation on the small ribosomal subunit. *Sci. Adv.* **7**, eabf7547 (2021).

24. Singh, S., Vanden Broeck, A., Miller, L., Chaker-Margot, M. & Klinge, S. Nucleolar maturation of the human small subunit processome. *Science* **373**, eabj5338 (2021).

25. Sykes, M. T. & Williamson, J. R. A complex assembly landscape for the 30S ribosomal subunit. *Annu. Rev. Biophys.* **38**, 197–215 (2009).

26. He, J. et al. Human C4orf14 interacts with the mitochondrial nucleoid and is involved in the biogenesis of the small mitochondrial ribosomal subunit. *Nucleic Acids Res.* **40**, 6097–6108 (2012).

27. Kolanczyk, M. et al. NOA1 is an essential GTPase required for mitochondrial protein synthesis. *Mol. Biol. Cell* **22**, 1–11 (2011).

28. Dennerlein, S., Rozanska, A., Wydro, M., Chrzanowska-Lightowlers, Z. M. A. & Lightowlers, R. N. Human ERAL1 is a mitochondrial RNA chaperone involved in the assembly of the 28S small mitochondrial ribosomal subunit. *Biochem. J.* **430**, 551–558 (2010).

29. Chatzispyrou, I. A. et al. A homozygous missense mutation in ERAL1, encoding a mitochondrial rRNA chaperone, causes Perrault syndrome. *Hum. Mol. Genet.* **26**, 2541–2550 (2017).

30. Liu, X. et al. Structural insights into dimethylation of 12S rRNA by TFB1M: indispensable role in translation of mitochondrial genes and mitochondrial function. *Nucleic Acids Res.* **47**, 7648–7665 (2019).

31. Rozanska, A. et al. The human RNA-binding protein RBFA promotes the maturation of the mitochondrial ribosome. *Biochem. J.* **474**, 2145–2158 (2017).

32. Gerber, S., Orssaud, C., Kaplan, J., Johansson, C. & Rozet, J.-M. MCAT mutations cause nuclear LHON-like optic neuropathy. *Genes (Basel)* **12**, 521 (2021).

33. Puchta, O. et al. DMR1 (CCM1/YGR150C) of *Saccharomyces cerevisiae* encodes an RNA-binding protein from the pentatricopeptide repeat family required for the maintenance of the mitochondrial 15S ribosomal RNA. *Genetics* **184**, 959–973 (2010).

34. Paul, M.-F., Alushin, G. M., Barros, M. H., Rak, M. & Tzagoloff, A. The putative GTPase encoded by MTG3 functions in a novel pathway for regulating assembly of the small subunit of yeast mitochondrial ribosomes. *J. Biol. Chem.* **287**, 24346–24355 (2012).

35. Hillen, H. S. et al. The pentatricopeptide repeat protein Rmd9 recognizes the dodecameric element in the 3′-UTRs of yeast mitochondrial mRNAs. *Proc. Natl Acad. Sci. USA* **118**, e2009329118 (2021).

36. Klinge, S. & Woolford, J. L. Ribosome assembly coming into focus. *Nat. Rev. Mol. Cell Biol.* **20**, 116–131 (2019).

37. Britton, R. A. Role of GTPases in bacterial ribosome assembly. *Annu. Rev. Microbiol.* **63**, 155–176 (2009).

38. Sudhamsu, J., Lee, G. I., Klessig, D. F. & Crane, B. R. The structure of YqeH. An AtNOS1/AtNOA1 ortholog that couples GTP hydrolysis to molecular recognition. *J. Biol. Chem.* **283**, 32968–32976 (2008).

39. Stephan, N. C., Ries, A. B., Boehringer, D. & Ban, N. Structural basis of successive adenosine modifications by the conserved ribosomal methyltransferase KsgA. *Nucleic Acids Res.* **49**, 6389–6398 (2021).

40. Aibara, S., Singh, V., Modelska, A. & Amunts, A. Structural basis of mitochondrial translation. *eLife* **9**, e58362 (2020).

41. Razi, A. et al. Role of Era in assembly and homeostasis of the ribosomal small subunit. *Nucleic Acids Res.* **47**, 8301–8317 (2019).

42. Nouet, C. et al. Rmd9p controls the processing/stability of mitochondrial mRNAs and its overexpression compensates for a partial deficiency of oxa1p in *Saccharomyces cerevisiae*. *Genetics* **175**, 1105–1115 (2007).

43. Herbert, C. J., Golik, P. & Bonnefoy, N. Yeast PPR proteins, watchdogs of mitochondrial gene expression. *RNA Biol.* **10**, 1477–1494 (2013).

44. Hunziker, M. et al. Conformational switches control early maturation of the eukaryotic small ribosomal subunit. *eLife* **8**, e45185 (2019).

45. Wiesenberger, G. & Fox, T. D. Pet127p, a membrane-associated protein involved in stability and processing of *Saccharomyces cerevisiae* mitochondrial RNAs. *Mol. Cell. Biol.* **17**, 2816–2824 (1997).

46. Replogle, J. M. et al. Mapping information-rich genotype–phenotype landscapes with genome-scale Perturb-seq. *Cell* **185**, 2559–2575.e28 (2022).

47. Shi, Z. et al. Mettl17, a regulator of mitochondrial ribosomal RNA modifications, is required for the translation of mitochondrial coding genes. *FASEB J.* **33**, 13040–13050 (2019).

48. Mireku, S. A., Sauer, M. M., Glockshuber, R. & Locher, K. P. Structural basis of nanobody-mediated blocking of BtuF, the cognate substrate-binding protein of the *Escherichia coli* vitamin B12 transporter BtuCD. *Sci. Rep.* **7**, 14296 (2017).

49. Gray, E. R. et al. Unravelling the molecular basis of high affinity nanobodies against HIV p24: in vitro functional, structural, and in silico insights. *ACS Infect. Dis.* **3**, 479–491 (2017).

50. Slaymaker, I. M. et al. Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2016).

51. Concordet, J.-P. & Haeussler, M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res.* **46**, W242–W245 (2018).

52. Clayton, D. & Shadel, G. S. Isolation of mitochondria from tissue culture cells. *Cold Spring Harb. Protoc.* **2014**, pdb.prot080002 (2014).

53. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).

54. Zivanov, J., Nakane, T. & Scheres, S. H. W. Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. *IUCrJ* **7**, 253–267 (2020).

55. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

56. Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).

57. Wagner, T. et al. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 218 (2019).

58. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

59. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).

60. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).

61. Tan, Y. Z. et al. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **14**, 793–796 (2017).

62. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

63. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).

64. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

65. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906 (2007).

66. Cox, J. et al. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell Proteomics* **13**, 2513–2526 (2014).

67. Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).

68. Itoh, Y., Naschberger, A., Mortezaei, N., Herrmann, J. M. & Amunts, A. Analysis of translating mitoribosome reveals functional characteristics of translation in mitochondria of fungi. *Nat. Commun.* **11**, 5187 (2020).

69. Khawaja, A. et al. Distinct pre-initiation steps in human mitochondrial translation. *Nat. Commun.* **11**, 2932 (2020).

Acknowledgements

We thank M. Ebrahim, J. Sotiris and H. Ng at the Evelyn Gruss Lipper Cryo-Electron Microscopy Resource Center at The Rockefeller University for assistance with grid screening and data collection; S. Mazel at the Flow Cytometry Resource Center at The Rockefeller University for assistance with single-cell sorting during human cell line generation; A. Vanden Broeck for advice with cryo-EM data processing and scripts for figures; and members of the Klinge laboratory for critical reading of this manuscript. MS data were generated by the Proteomics Resource Center at The Rockefeller University (RRID: SCR_017797) using instrumentation funded by the Sohn Conferences Foundation and the Loena M. and Harry B. Helmsley Charitable Trust, and we particularly thank H. Molina, S. Heissel and C. Peralta for assistance. This study was funded, in part, by the NIH T32 training grant 5T32GM136640-02 (to N.J.H.). and by the Robertson Foundation (to S.K.).

Author information

Author notes

1. These authors contributed equally: Nathan J. Harper, Chloe Burnside

Authors and Affiliations

1. Laboratory of Protein and Nucleic Acid Chemistry, The Rockefeller University, New York, NY, USA

   Nathan J. Harper, Chloe Burnside & Sebastian Klinge

2. Tri-Institutional Training Program in Chemical Biology, The Rockefeller University, New York, NY, USA

   Nathan J. Harper & Chloe Burnside

Contributions

S.K. conceived the study. N.J.H., C.B. and S.K. designed the experiments and analysed the data. N.J.H. performed all experiments in human cells, including cell line generation, complex purification, cryo-EM data processing and model building. C.B. performed all experiments in yeast cells, including strain generation, complex purification, cryo-EM data processing, model building and northern blotting. All authors wrote and edited the manuscript.

Corresponding author

Correspondence to Sebastian Klinge.

Ethics declarations

Competing interests

The Rockefeller University has filed a patent related to the human genome editing platform used in this study, on which S.K. is named as an inventor. All other authors declare no competing interests.

Peer review

Peer review information

Nature thanks Joaquin Ortega and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Roles of evolutionarily unique mitoribosomal proteins and assembly factors in mitochondrial small subunit biogenesis.

Models of: (**a**) human mitoribosomal proteins not present in yeast in context of the mature subunit (PDB:3J9M)[21] (**b**) yeast mitoribosomal proteins and rRNA not present in humans in context of the mature subunit (PDB:5MRC)[19]. Unique rRNA segments are colored white unless the region is functionally relevant within assembly states captured in this study. (**c**) human assembly factors identified in this study. (**d**) yeast assembly factors identified in this study. For panels a-d, bottom images include boxes outlining labeled regions of interest for panels e-h. (**e**) the head domain near METTL17/Rsm22 in States D and E (human) and State 3 and the mature structure (yeast, PDB:5MRC)[19]. In humans, unique protein mS31 binds unique protein mS39. In yeast, a unique rRNA segment, h33, is accommodated by unique protein mS46 after Rsm22 dissociation. Human assembly factor MCAT, which lacks a yeast homologue, may be functionally replaced by yeast uS13m and uS19m. (**f**) proteins accommodating the 5′ rRNA in State A (human) and states 2 and 3 (yeast). The core conserved binding module formed by uS5m is augmented in yeast to integrate the 5′ precursor and Ccm1 via yeast specific proteins uS4m, mS45, and mS47. (**g**) proteins involved in 3′ stabilization in States B and C (human) and State 3 (yeast). In humans, dedicated 3′ binding assembly factors ERAL1 and RBFA stabilize the 3′ rRNA during assembly. In yeast, the 3′ rRNA is extended and is accommodated via a backpack module including unique proteins mS42,

mS43, and an mS23 extension. (**h**) proteins involved in h44 docking in States A and B (human) and State 3 (yeast). Human specific proteins mS27 and mS34 allow NOA1-mediated inhibition of h44 docking, while h44 in yeast is bound by yeast protein mS44. For panels e-h, conserved proteins and rRNA segments between the two assembly systems are colored identically.

Extended Data Fig. 2 Details of NOA1 function within State A.

(**a**) Interaction network of NOA1 in State A in context with the entire complex (left), shown in cartoon (right), and in a schematic (bottom). (**b**) Model and cryo-EM map of region highlighted in green in (a). (**c**) Conservation analysis of NOA1 N-terminal peptide region (left) and molecular context, shown as model and cryo-EM density. (**d**) Models of the GTPase domain of related circularly permutated GTPase YqeH (left, PDB:3EC1)[38] compared to NOA1 (right). Conserved GTPase motifs are colored and labeled. Proposed rearrangements of NOA1 occurring during GTP binding are shown as arrows. Conservation analysis of the helical insertion in the GTPase domain (right).

Extended Data Fig. 3 TFB1M plays roles in h44 maturation and h45 methylation.

(**a**) Interaction network of TFB1M in State C in context with the entire complex (left), shown in cartoon (right) and in a schematic (bottom). (**b**) Models and cryo-EM maps showing TFB1M h44 clamping activity. (**c**) Models and cryo-EM maps showing the enzymatic active site lacking density for methyl donor S-adenosyl methionine. (**d**) Comparison of h45 binding mode between TFB1M and its bacterial homologue KsgA (PDB:7O5H)[39]. (**e**) Comparison of h45 cryo-EM density between State C and State D, showing conformational changes and appearance of density corresponding to modified adenosine residues. Maps from both states are displayed at the same contour level.

Extended Data Fig. 4 Conserved and distinct features of decoding center maturation in *S. cerevisiae*.

(**a**) Overall structure of yeast State 3. Black and red squares denote location of panels c and d, and g-j, respectively. (**b**) Overall structure of human State E. The black square denotes location of panels e and f. (**c**) Model of h31 bound by Rsm22, as observed in State 3. The grey arrow denotes relative movement of rRNA towards the decoding center due to subunit compaction after Rsm22 dissociation. (**d**) Same view as (c) in the structure of the *N. crassa* mitoribosome during translation (PDB:6YWY)[68], showing interactions of h31 with mRNA and the P-site tRNA. (**e**) Model of h31 bound by METTL17, as observed in State E. The grey arrow denotes relative movement of rRNA towards the decoding center due to subunit compaction after METTL17 dissociation. (**f**) Same view as (e) in the structure of the *H. sapiens* mitoribosome during translation bound to A, P, and E-site tRNAs (PDB:6ZSG)[40], showing interactions of h31 with mRNA and the P-site tRNA. In panels c-f, models are aligned on h31. (**g**–**j**) Timeline of yeast States 1–3 and the mature structure (PDB:5MRC)[19] with associated map gaussian filtered at 1.4 standard deviations overlaid. Note absence of h45 in State 1 and mS38 in assembly intermediates States 1–3.

Extended Data Fig. 5 Structural and mechanistic context of assembly factor mutations implicated in disease.

(**a**) Model of State B highlighting the context of ERAL1 during assembly. (**b**) Location of the N236I mutant implicated in Perrault syndrome in the GTPase binding site of ERAL1[29]. (**c**) Same view as (b) in State C, showing structural transitions associated with ERAL1/RBFA exchange. (**d**) Model of State D highlighting the context of MCAT during assembly. (**e**) Location of MCAT L81R, R212W, and E347K mutations implicated in LHON-like optic neuropathy[32]. (**f**) Same view as (e) in State E, showing structural transitions associated with MCAT dissociation. In panels b, c, e, and f, properties of local rRNA elements within each state are noted.

Extended Data Fig. 6 The 3′ rRNA binding mode of RBFA is distinct from that of *E. coli*.

(**a**) Interaction network of RBFA in State C in context with the entire complex (left), shown in cartoon (right) and in a schematic (bottom). (**b**)

Model and cryo-EM map of the highlighted region in (a), showing the composite binding site formed by RBFA and bS21m. (**c**) Conservation analysis of the N-terminal peptide segment observed to interact with bS21m during 3′ rRNA binding. (**d**) Model illustrating conformational changes in uS7m during RBFA/ERAL1 exchange in States B and C. (**e**) Comparison of the 3′ rRNA binding mode between human and bacterial systems (PDB:7BOH)[23].

Extended Data Fig. 7 Molecular details of NAD binding and platform stabilization.

(**a**,**b**) rRNA elements involved in platform stabilization during the transition between state B (panel a) and state C (panel b), shown in context with the rest of the complex (left), and as cartoons (right). Movement of rRNA associated with platform stabilization shown with arrows. (**c**,**d**) Models and cryo-EM density showing transition of platform rRNA between state B (panel c) and state C (panel d). Arrows in all panels highlight conformational changes occurring during the transition from state B to state C.

Extended Data Fig. 8 A conserved role of METTL17/Rsm22 in head maturation.

(**a**–**c**) Models of METTL17, Rsm22, and SAF1 (Rsm22 in *T. brucei*, PDB:6SGB)[15] respectively, with associated rRNA elements (top). Below, cryo-EM densities and models for ligands and their binding sites associated with each assembly factor (for *T. brucei*, map is EMD-10180, and empty density was originally modeled as a Zn^{2+})[15]. (**d**,**e**) Ordering of yeast mitoribosomal protein elements (shown as cartoon) in the head domain induced upon Rsm22 dissociation and head compaction. Peptides which become ordered between State 3 and the mature structure are colored in a darker shade. Additional protein and RNA components are shown as transparent surfaces. Mature structure derived from PDB:5MRC[19]. Models of RNA omitted for clarity. (**f**,**g**) Ordering of human mitoribosomal protein elements (shown as cartoon) in the head domain induced upon METTL17 dissociation and head compaction. Peptides which become ordered between

State E and the mature structure are colored in a darker shade. Additional protein and RNA components are shown as transparent surfaces. Mature structure derived from PDB:6RW4[69]. Previously ordered RNA omitted for clarity in mature structures.

Extended Data Table 1 Cryo-EM data collection and refinement statistics of human structures

Extended Data Table 2 Cryo-EM data collection and refinement statistics of yeast structures

Supplementary information

Supplementary Figures

Supplementary Figs. 1–27.

Reporting Summary

Supplementary Data 1–5

Supplementary Data files 1–4 contain identified peptides in LC-MS/MS human METTL1, LC-MS/MS yeast Ccm1, LC-MS/MS yeast Rsm22 and LC-MS/MS yeast Mtg3 samples. Supplementary Data 5 contains comparative mass spec analysis of WT and PET127 KO yeast mtSSU particles isolated via uS17m. IP-MS/MS experiments were performed in triplicate for each condition (n = 3 biologically independent experiments). A 2-sided student t-test was used to determine p-values for changes in protein abundance between WT and PET127 KO mtSSU purification experiments.

PyMOL Session Files

StateA.pse PyMOL session including the structure of human State A with annotated proteins and rRNA, coloured as in Fig. 1. **StateB.pse** PyMOL session including the structure of human State B with annotated proteins and rRNA, coloured as in Fig. 1. **StateC.pse** PyMOL session including the structure of human State C with annotated proteins and rRNA, coloured as

in Fig. 1. **StateD.pse** PyMOL session including the structure of human State D with annotated proteins and rRNA, coloured as in Fig. 1. **StateE.pse** PyMOL session including the structure of human State E with annotated proteins and rRNA, coloured as in Fig. 1. **StateCstar.pse** PyMOL session including the structure of human State C* with annotated proteins and rRNA, coloured as in Fig. 1. **State1.pse** PyMOL session including the structure of yeast State 1 with annotated proteins and rRNA, coloured as in Fig. 1. **State2.pse** PyMOL session including the structure of yeast State 2 with annotated proteins and rRNA, coloured as in Fig. 1. **State3.pse** PyMOL session including the structure of yeast State 3 with annotated proteins and rRNA, coloured as in Fig. 1. **comparative_human_states.pse** PyMOL session including the aligned structures of human States A-E, each with annotated proteins and rRNA, coloured as in Fig. 1. **comparative_yeast_states.pse** PyMOL session including the aligned structures of yeast States 1-3, each with annotated proteins and rRNA, coloured as in Fig. 1.

Peer Review File

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Harper, N.J., Burnside, C. & Klinge, S. Principles of mitoribosomal small subunit assembly in eukaryotes. *Nature* **614**, 175–181 (2023). https://doi.org/10.1038/s41586-022-05621-0

- Received: 13 July 2022
- Accepted: 02 December 2022
- Published: 08 December 2022
- Issue Date: 02 February 2023
- DOI: https://doi.org/10.1038/s41586-022-05621-0

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05621-0

Amendments & Corrections

- **Author Correction: Cyclic nucleotide-induced helical structure activates a TIR immune effector** [11 Jan 2023]
  -
- **Author Correction: NLR surveillance of pathogen interference with hormone receptors induces immunity** [11 Jan 2023]
  -
- **Author Correction: Global hotspots of salt marsh change and carbon emissions** [18 Jan 2023]
  -
- **Publisher Correction: Structured cerebellar connectivity supports resilient pattern separation** [11 Jan 2023]
  -
- **Publisher Correction: Cost-effective mitigation of nitrogen pollution from global croplands** [16 Jan 2023]
  -

- Author Correction
- Published: 11 January 2023

Author Correction: Cyclic nucleotide-induced helical structure activates a TIR immune effector

- Gaëlle Hogrel,
- Abbie Guild,
- Shirley Graham,
- Hannah Rickman,
- Sabine Grüschow,
- Quentin Bertrand,
- Laura Spagnolo &
- …
- Malcolm F. White

Show authors

Nature **volume 614**, page E15 (2023)

- 369 Accesses
- 1 Altmetric
- Metrics details

The Original Article was published on 10 August 2022

Correction to: *Nature* https://doi.org/10.1038/s41586-022-05070-9
Published online 10 August 2022

In the version of this article initially published, the Acknowledgements section omitted thanks to the European Synchrotron Radiation Facility for provision of synchrotron radiation facilities, which has now been included in the HTML and PDF versions of the article.

Author information

Author notes

1. Quentin Bertrand

   Present address: Laboratory of Biomolecular Research, Biology and Chemistry Division, Paul Scherrer Institute, Villigen, Switzerland

Authors and Affiliations

1. School of Biology, University of St Andrews, St Andrews, UK

   Gaëlle Hogrel, Shirley Graham, Hannah Rickman, Sabine Grüschow & Malcolm F. White

2. Institute of Molecular, Cell and Systems Biology, University of Glasgow, Glasgow, UK

   Abbie Guild & Laura Spagnolo

3. Centre for Regenerative Medicine, Institute for Regeneration and Repair, The University of Edinburgh, Edinburgh, UK

   Abbie Guild

4. Université Grenoble Alpes, CNRS, CEA, IBS, Grenoble, France

   Quentin Bertrand

Corresponding authors

Correspondence to Laura Spagnolo or Malcolm F. White.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Hogrel, G., Guild, A., Graham, S. *et al.* Author Correction: Cyclic nucleotide-induced helical structure activates a TIR immune effector. *Nature* **614**, E15 (2023). https://doi.org/10.1038/s41586-023-05701-9

- Published: 11 January 2023
- Issue Date: 02 February 2023
- DOI: https://doi.org/10.1038/s41586-023-05701-9

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05701-9

- Author Correction
- Published: 11 January 2023

Author Correction: NLR surveillance of pathogen interference with hormone receptors induces immunity

- Jing Chen,
- Yanxiao Zhao,
- Xuanjie Luo,
- Hao Hong,
- Tongqing Yang,
- Shen Huang,
- Chunli Wang,
- Hongyu Chen,
- Xin Qian,
- Mingfeng Feng,
- Zhengqiang Chen,
- Yongxin Dong,
- Zhenchuan Ma,
- Jia Li,
- Min Zhu,
- Sheng Yang He,
- Savithramma P. Dinesh-Kumar &
- …
- Xiaorong Tao

Show authors

Nature **volume 614**, page E16 (2023)

- 435 Accesses
- 1 Altmetric
- Metrics details

The Original Article was published on 14 December 2022

Correction to: *Nature* https://doi.org/10.1038/s41586-022-05529-9
Published online 14 December 2022

In the version of this article initially published, there was a spurious line blocking the arrow from the transcriptional repressor to the 26S proteasome drawn in the center panel of Fig. 4h. The line has now been removed from the HTML and PDF versions of the article.

Author information

Author notes

1. These authors contributed equally: Jing Chen, Yanxiao Zhao

Authors and Affiliations

1. Key Laboratory of Plant Immunity, Department of Plant Pathology, Nanjing Agricultural University, Nanjing, China

   Jing Chen, Yanxiao Zhao, Xuanjie Luo, Hao Hong, Tongqing Yang, Shen Huang, Chunli Wang, Hongyu Chen, Xin Qian, Mingfeng Feng, Zhengqiang Chen, Yongxin Dong, Zhenchuan Ma, Jia Li, Min Zhu & Xiaorong Tao

2. Howard Hughes Medical Institute, Department of Biology, Duke University, Durham, NC, USA

   Sheng Yang He

3. Department of Plant Biology and The Genome Center, College of Biological Sciences, University of California, Davis, CA, USA

Savithramma P. Dinesh-Kumar

Corresponding author

Correspondence to Xiaorong Tao.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Chen, J., Zhao, Y., Luo, X. *et al.* Author Correction: NLR surveillance of pathogen interference with hormone receptors induces immunity. *Nature* **614**, E16 (2023). https://doi.org/10.1038/s41586-023-05702-8

- Published: 11 January 2023
- Issue Date: 02 February 2023
- DOI: https://doi.org/10.1038/s41586-023-05702-8

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05702-8

- Author Correction
- Open Access
- Published: 18 January 2023

Author Correction: Global hotspots of salt marsh change and carbon emissions

- Anthony D. Campbell ,
- Lola Fatoyinbo,
- Liza Goldberg &
- …
- David Lagomasino

Show authors

Nature **volume 614**, page E17 (2023)

- 303 Accesses

- Metrics details

The Original Article was published on 30 November 2022

Correction to: *Nature* https://doi.org/10.1038/s41586-022-05355-z
Published online 30 November 2022

In the version of this article initially published, in the fourth sentence of the ‘Hotspots of salt marsh change’ section, the estimate of the global loss rate of salt marsh should have read 0.28% year^{-1}, not 0.26%. In addition, the Acknowledgements section and the Fig. 3 caption didn’t note that Maxar

data were provided under the National Geospatial-Intelligence Agency’s NextView license agreement. The errors have been corrected in the HTML and PDF versions of the article

Author information

Authors and Affiliations

1. Biospheric Sciences Laboratory, National Aeronautics and Space Administration (NASA) Goddard Space Flight Center, Greenbelt, MD, USA

   Anthony D. Campbell, Lola Fatoyinbo & Liza Goldberg

2. NASA Postdoctoral Program, Oak Ridge Associated Universities, Oak Ridge, TN, USA

   Anthony D. Campbell

3. GESTAR II, University of Maryland, Baltimore County, Baltimore, MD, USA

   Anthony D. Campbell

4. Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA

   Liza Goldberg

5. Integrated Coastal Programs, East Carolina University, Wanchese, NC, USA

   David Lagomasino

Corresponding author

Correspondence to Anthony D. Campbell.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Campbell, A.D., Fatoyinbo, L., Goldberg, L. *et al.* Author Correction: Global hotspots of salt marsh change and carbon emissions. *Nature* **614**, E17 (2023). https://doi.org/10.1038/s41586-023-05698-1

- Published: 18 January 2023
- Issue Date: 02 February 2023
- DOI: https://doi.org/10.1038/s41586-023-05698-1

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05698-1

- Publisher Correction
- Published: 11 January 2023

Publisher Correction: Structured cerebellar connectivity supports resilient pattern separation

- Tri M. Nguyen,
- Logan A. Thomas,
- Jeff L. Rhoades,
- Ilaria Ricchi,
- Xintong Cindy Yuan,
- Arlo Sheridan,
- David G. C. Hildebrand,
- Jan Funke,
- Wade G. Regehr &
- …
- Wei-Chung Allen Lee

Show authors

Nature **volume 614**, page E18 (2023)

- 330 Accesses
- 1 Altmetric
- Metrics details

The Original Article was published on 23 November 2022

Correction to: *Nature* https://doi.org/10.1038/s41586-022-05471-w
Published online 23 November 2022

In the version of this article initially published, in the “MF bouton locations” graph of Fig. 2b, the 100 µm tick mark on the *y* axis was shifted down and therefore displayed the incorrect value. Furthermore, the input similarity value in Fig. 3b, now reading “6/9 (66.7%),” was shown as “5/9 (55.5%).” The errors have been corrected in the HTML and PDF versions of the article.

Author information

Author notes

1. Logan A. Thomas

   Present address: Biophysics Graduate Group, University of California Berkeley, Berkeley, CA, USA

2. Ilaria Ricchi

   Present address: Neuro-X Institute, École Polytechnique Fédérale de Lausanne (EPFL), Geneva, Switzerland

3. Arlo Sheridan

   Present address: Waitt Advanced Biophotonics Center, Salk Institute for Biological Studies, La Jolla, CA, USA

4. David G. C. Hildebrand

   Present address: Laboratory of Neural Systems, The Rockefeller University, New York, NY, USA

5. These authors contributed equally: Tri M. Nguyen, Logan A. Thomas

Authors and Affiliations

1. Department of Neurobiology, Harvard Medical School, Boston, MA, USA

   Tri M. Nguyen, Logan A. Thomas, Jeff L. Rhoades, Ilaria Ricchi, Xintong Cindy Yuan, David G. C. Hildebrand & Wade G. Regehr

2. Program in Neuroscience, Division of Medical Sciences, Graduate School of Arts and Sciences, Harvard University, Cambridge, MA, USA

   Jeff L. Rhoades & Xintong Cindy Yuan

3. HHMI Janelia Research Campus, Ashburn, VA, USA

   Arlo Sheridan & Jan Funke

4. F. M. Kirby Neurobiology Center, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA

   Wei-Chung Allen Lee

Corresponding author

Correspondence to Wei-Chung Allen Lee.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Nguyen, T.M., Thomas, L.A., Rhoades, J.L. *et al.* Publisher Correction: Structured cerebellar connectivity supports resilient pattern separation. *Nature* **614**, E18 (2023). https://doi.org/10.1038/s41586-023-05703-7

- Published: 11 January 2023
- Issue Date: 02 February 2023
- DOI: https://doi.org/10.1038/s41586-023-05703-7

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05703-7

- Publisher Correction
- Open Access
- Published: 16 January 2023

Publisher Correction: Cost-effective mitigation of nitrogen pollution from global croplands

- Baojing Gu ,
- Xiuming Zhang,
- Shu Kee Lam,
- Yingliang Yu,
- Hans J. M. van Grinsven,
- Shaohui Zhang,
- Xiaoxi Wang,
- Benjamin Leon Bodirsky,
- Sitong Wang,
- Jiakun Duan,
- Chenchen Ren,
- Lex Bouwman,
- Wim de Vries,
- Jianming Xu ,
- Mark A. Sutton &
- …
- Deli Chen

Show authors

Nature **volume 614**, page E19 (2023)

- 451 Accesses

- 1 Altmetric
- Metrics details

The Original Article was published on 04 January 2023

Correction to: *Nature* https://doi.org/10.1038/s41586-022-05481-8
Published online 4 January 2023

This paper was originally published under a standard Springer Nature license (© The Author(s), under exclusive licence to Springer Nature Limited). It is now available as an Open Access paper under a Creative Commons Attribution 4.0 International license, © The Author(s). Additionally, the maps in Figure 3d,e were mistakenly reversed (e.g., Alaska should now appear in red in the top-left of Fig. 3d), while in Figure 5c, the y-axis label for cropland area read "10^3 ha" rather than "10^6 ha" as corrected now. The errors have been corrected in the HTML and PDF versions of the article and the source data for are updated.

Author information

Authors and Affiliations

1. College of Environmental and Resource Sciences, Zhejiang University, Hangzhou, China

   Baojing Gu, Sitong Wang, Jiakun Duan & Jianming Xu

2. Policy Simulation Laboratory, Zhejiang University, Hangzhou, China

   Baojing Gu, Sitong Wang, Jiakun Duan & Chenchen Ren

3. School of Agriculture and Food, The University of Melbourne, Melbourne, Victoria, Australia

   Xiuming Zhang, Shu Kee Lam & Deli Chen

4. Key Laboratory of Agricultural Environment of the Lower Reaches of the Yangtze River, Institute of Agricultural Resources and Environment, Jiangsu Academy of Agricultural Sciences, Nanjing, China

   Yingliang Yu

5. PBL Netherlands Environmental Assessment Agency, The Hague, The Netherlands

   Hans J. M. van Grinsven & Lex Bouwman

6. School of Economics and Management, Beihang University, Beijing, China

   Shaohui Zhang

7. International Institute for Applied Systems Analysis, Laxenburg, Austria

   Shaohui Zhang

8. China Academy for Rural Development, Zhejiang University, Hangzhou, China

   Xiaoxi Wang

9. Department of Agricultural Economics and Management, School of Public Affairs, Zhejiang University, Hangzhou, China

   Xiaoxi Wang

10. Potsdam Institute for Climate Impact Research (PIK), Potsdam, Germany

    Xiaoxi Wang & Benjamin Leon Bodirsky

11. Department of Earth Sciences - Geochemistry, Faculty of Geosciences, Utrecht University, Utrecht, The Netherlands

Lex Bouwman

12. Environmental Systems Analysis Group, Wageningen University & Research, Wageningen, The Netherlands

    Wim de Vries

13. Zhejiang Provincial Key Laboratory of Agricultural Resources and Environment, Zhejiang University, Hangzhou, China

    Jianming Xu

14. Edinburgh Research Station, UK Centre for Ecology & Hydrology, Penicuik, UK

    Mark A. Sutton

Corresponding authors

Correspondence to Baojing Gu or Jianming Xu.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Gu, B., Zhang, X., Lam, S.K. *et al.* Publisher Correction: Cost-effective mitigation of nitrogen pollution from global croplands. *Nature* **614**, E19 (2023). https://doi.org/10.1038/s41586-023-05708-2

- Published: 16 January 2023
- Issue Date: 02 February 2023
- DOI: https://doi.org/10.1038/s41586-023-05708-2

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05708-2